

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 0.461066 Seconds
(without alignments)
1043.144 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.78.*

1: Pirl.*

2: Pirl2.*

3: Pirl3.*

4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	86	2	cell division topo
2	21	100.0	86	2	cell division topo
3	21	100.0	132	2	hypothetical prote
4	21	100.0	171	2	hypothetical prote
5	21	100.0	181	2	hypothetical prote
6	21	100.0	188	2	phosphoribosylglyc
7	21	100.0	194	2	hypothetical prote
8	21	100.0	197	2	hypothetical prote
9	21	100.0	200	2	probable sugar-pho
10	21	100.0	203	2	hypothetical prote
11	21	100.0	203	2	hypothetical prote
12	21	100.0	219	2	tetr family bacter
13	21	100.0	219	2	transcription regu
14	21	100.0	224	1	ULI protein - huma
15	21	100.0	236	2	basal-body rod mod
16	21	100.0	243	2	probable membrane-
17	21	100.0	250	1	nonstructural prot
18	21	100.0	263	1	probable S-adenosy
19	21	100.0	266	2	conserved hypothet
20	21	100.0	268	2	ribosomal protein
21	21	100.0	271	1	pancreatic elastas
22	21	100.0	279	2	probable 1-acylgly
23	21	100.0	282	1	cytochrome c1 - Rh
24	21	100.0	293	2	mating type protei
25	21	100.0	301	2	probable Glycosylt
26	21	100.0	305	2	acetyltransferase
27	21	100.0	331	2	hypothetical prote
28	21	100.0	335	2	DNA-directed DNA p
29	21	100.0	335	2	DNA-directed DNA p

ALIGNMENTS

RESULT 1

AI2955

cell division topological specificity factor minE [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AI2955

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AI2955

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAU44063.1; PID:gl7741627; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: minE

A:Map position: linear chromosome

Query Match 100.0%; Score 21; DB 2; Length 86;

Best Local Similarity 100.0%; Pred.No. 55;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5

Db 45 SVIAK 49

RESULT 2

D98327

cell division topological specificity factor [imported] - Agrobacterium tumefaciens (str
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: D98327

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: D98327

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK90142.1; PID:gl5160139; GSPDB:GN00170

C:Genetics:

A:Gene: AGR_L 3134

A:Map position: linear chromosome

aquaporin 9 - huma
transforming prote
early switch prote
probable hexosyltr
probable dioxysena
probable MYB famil
hypothetical prote
hypothetical prote
early switch prote
hypothetical prote
hypothetical prote
early switch prote
hypothetical prote
hypothetical prote
hypothetical prote

30 21 100.0 342 2 JCS791
31 21 100.0 348 2 S11198
32 21 100.0 355 1 A55473
33 21 100.0 357 1 G69290
34 21 100.0 358 2 C84713
35 21 100.0 367 2 T01017
36 21 100.0 396 2 B96601
37 21 100.0 399 2 T49186
38 21 100.0 417 2 B55473
39 21 100.0 418 2 A96601
40 21 100.0 420 2 AF2464
41 21 100.0 426 2 C55473
42 21 100.0 430 2 S15308
43 21 100.0 470 2 G85877
44 21 100.0 470 2 F91033
45 21 100.0 498 2 T30092

Query Match 100.0%; Score 21; DB 2; Length 86;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 DB 45 SVIAK 49

RESULT 3

T32373

hypothetical protein C01B12.7 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000

C;Accession: T32373

R;Sheet, P.; Maggii, L.

submitted to the EMBL Data Library, September 1997

A;Description: The sequence of C. elegans cosmid C01B12.

A;Reference number: Z21156

A;Accession: T32373

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-132 <SCH>

A;Cross-references: EMBL:AF025458; PIDN:AAB70973.1; GSPDB:GN000020; CESP:C01B12.7

A;Experimental source: strain Bristol N2; clone C01B12

C;Genetics:

A;Gene: CESP:C01B12.7

A;Map position: 2

A;Introns: 23/3; 90/2

C;Superfamily: Caenorhabditis elegans hypothetical protein C01B12.7

Query Match 100.0%; Score 21; DB 2; Length 132;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 DB 25 SVIAK 29

RESULT 4

G90532

hypothetical protein MYPU_1670 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C;Species: Mycoplasma pulmonis

C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C;Accession: G90532

R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm

A;Reference number: A99512; MUID:21267165; PMID:11353084

A;Accession: G90532

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-171 <KUR>

A;Cross-references: GB:AL445566; PID:gl14089580; PIDN:CAC13340.1; GSPDB:GN00153

A;Experimental source: strain UAB CTIP

C;Genetics:

A;Gene: MYPU_1670

A;Genetic code: SGC3

Query Match 100.0%; Score 21; DB 2; Length 171;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 DB 117 SVIAK 121

RESULT 5

AF1931

hypothetical protein alr1001 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AF1931

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi,

Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.;

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AF1931

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-181 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA872958.1; PID:gl7130347; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr1001

Query Match 100.0%; Score 21; DB 2; Length 181;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 DB 37 SVIAK 41

RESULT 6

F89876

phosphoribosylglycinamide formyltransferase [imported] - Staphylococcus aureus (strain N

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 21-Jun-2002

C;Accession: F89876

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mitutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: F89876

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-188 <KUR>

A;Cross-references: GB:BA000018; PID:gl3700873; PIDN:BA842169.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: purN

Query Match 100.0%; Score 21; DB 2; Length 188;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 DB 181 SVIAK 185

RESULT 7

T15115

hypothetical protein ZC132.9 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C;Accession: T15115

R;Bradshaw, H.; Devlin, K.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of C. elegans cosmid ZC132.

A;Reference number: Z18294

A;Accession: T15115

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-194 <BRA>

A;Cross-references: EMBL:AF014939; MID:g2275620; PID:g2275628; PIDN:AAB63931.1; GSPDB:GN0

A;Experimental source: strain Bristol N2; clone ZC132

C:Genetics:
A:Gene: CESP:ZC132.9
A:Map position: 5
A:Introns: 135/3; 153/3

Query Match 100.0%; Score 21; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
|||||
Db 56 SVIAK 60

RESULT 8

D87309
hypothetical protein CC0485 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: D87309
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Koln
n, J.; Molaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: D87309

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-197 <STO>

A:CROSS-references: GB:AE005673; NID:gl3421662; PIDN:AAK22472.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0485

C:Superfamily: 50S ribosomal protein L25

Query Match 100.0%; Score 21; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
|||||
Db 73 SVIAK 77

RESULT 9

AG1137
probable sugar-phosphate isomerase homolog lmc0502 [imported] - Listeria monocytogenes
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG1137
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative Genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1137

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-200 <GLA>

A:CROSS-references: GB:NC_003210; PIDN:CAC98581.1; PID:gl6409878; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmc0502

Query Match 100.0%; Score 21; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
|||||

Db 130 SVIAK 134

RESULT 10

C85288

hypothetical protein AT4g24980 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: C85288

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: C85288

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-203 <STO>

A:CROSS-references: GB:NC_001268; NID:g7269348; PIDN:CAB79407.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g24980

A:Map position: 4

Query Match 100.0%; Score 21; DB 2; Length 203;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5

|||||

Db 30 SVIAK 34

RESULT 11

T05519

hypothetical protein F13M23.120 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999

C:Accession: T05519

R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.J

submitted to the Protein Sequence Database, February 1999

A:Reference number: 215419

A:Accession: T05519

A:Molecule type: DNA

A:Residues: 1-203 <BEV>

A:CROSS-references: EMBL:AL035523

A:Experimental source: cultivar Columbia; BAC clone F13M23

C:Genetics:

A:Map position: 4

A>Note: intron positions not resolved

A:Note: F13M23.120

Query Match 100.0%; Score 21; DB 2; Length 203;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5

|||||

Db 30 SVIAK 34

RESULT 12

A97665

tetR family bacterial regulatory protein (AF232237) [imported] - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C:Accession: A97665

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: A97665

A>Status: preliminary

A:Molecule type: DNA

A;Residues: 1-219 <KUR>
 A;Cross-references: GB:AE007869; PIDN:AAK88274.1; PID:g15157738; GSPDB:GN00169
 C;Genetics:
 A;Gene: AGR_C_4617
 A;Map position: circular chromosome

Query Match 100.0%; Score 21; DB 2; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SVIAK 5
 |||||
 Db 146 SVIAK 150

RESULT 13

AD2889
 transcription regulator, TetR family amrE [imported] - Agrobacterium tumefaciens (strain
 C;Species: Agrobacterium tumefaciens
 C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C;Accession: AD2889

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AD2889

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-219 <KUR>

A;Cross-references: GB:AE008688; PIDN:AAL43530.1; PID:g17741041; GSPDB:GN00186

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: amrE

A;Map position: circular chromosome

Query Match 100.0%; Score 21; DB 2; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SVIAK 5
 |||||
 Db 146 SVIAK 150

RESULT 14

WMBEHG

UL1 protein - human herpesvirus 2 (strain HG52)

C;Species: human herpesvirus 2

A;Note: host Homo sapiens (man)

C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000

C;Accession: JQ1494

R;McGeoch, D.J.; Cunningham, C.; McIntyre, G.; Dolan, A.

J. Gen. Virol. 72, 3057-3075, 1991

A;Title: Comparative sequence analysis of the long repeat regions and adjoining parts of

A;Reference number: JQ1494; MUID:92113549; PMID:1662697

A;Accession: JQ1494

A;Molecule type: DNA

A;Residues: 1-224 <MCG>

A;Cross-references: GB:D10470; DBBJ:D01127; NID:G221791; PIDN:BAA01264.1; PID:G221792

C;Genetics:

A;Gene: UL1

C;Superfamily: varicella-zoster virus gene 60 protein

Query Match 100.0%; Score 21; DB 1; Length 224;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SVIAK 5
 |||||

Db 30 SVIAK 34

RESULT 15

D84969

basal-body rod modification protein flgD [imported] - Buchnera sp. (strain APS)
 C;Species: Buchnera sp.

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001

C;Accession: D84969

R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.

A;Reference number: AB4930; MUID:20445173; PMID:10993077

A;Accession: D84969

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-236 <STO>

A;Cross-references: GB:AP000398; GSPDB:GN00144

A;Experimental source: strain APS

C;Genetics:

A;Gene: flgD; BU339

Query Match 100.0%; Score 21; DB 2; Length 236;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SVIAK 5
 |||||
 Db 186 SVIAK 190

Search completed: August 12, 2004, 06:13:48
 Job time : 3.46107 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 0.307377 Seconds
(without alignments)
847.008 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLQSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	86	1 MINE AGRT5	Q8uax0 agrobacteri
2	21	100.0	95	1 RL27 MYCPE	Q8evw5 mycoplasma
3	21	100.0	101	1 SR19 METAC	Q8tyt3 methanosarc
4	21	100.0	101	1 SR19 METWA	Q8pwm7 methanosarc
5	21	100.0	134	1 YATB HAEIN	Q57425 haemophilus
6	21	100.0	188	1 PUR3 STAAW	Q99v25 staphylococ
7	21	100.0	188	1 PUR3 STAAW	Q8nx89 staphylococ
8	21	100.0	224	1 RNH2 SYNXP	RP
9	21	100.0	224	1 VGLL_HSV2H	P28278 herpes simp
10	21	100.0	225	1 RS3 THEAC	Q9h1r5 thermoplas
11	21	100.0	225	1 RS3 THEAC	Q97bx1 thermoplas
12	21	100.0	233	1 RNH2 STRAW	Q82kf0 streptomyce
13	21	100.0	236	1 FLGD BUCAI	P57421 buchnera ap
14	21	100.0	250	1 VNST_PTPV	P03516 punta toro
15	21	100.0	253	1 EXBB_XANCP	Q34260 xanthomonas
16	21	100.0	263	1 KSGA_MYCPN	P75113 m dimethyla
17	21	100.0	267	1 RS3 MYCPE	P47403 mycoplasma
18	21	100.0	271	1 EL2 RAT	P00774 rattus norv
19	21	100.0	282	1 CY1 RHQVI	P01379 rhodospoedu
20	21	100.0	293	1 MAYA_NUCR	P19392 neurospora
21	21	100.0	320	1 OYNI_HUMAN	Q8n53 homo sapien
22	21	100.0	331	1 Y542_RICCN	Q92178 rickettsia
23	21	100.0	333	1 DPOB_XENLA	O57363 xenopus lae
24	21	100.0	334	1 DPOB_HUMAN	P06746 homo sapien
25	21	100.0	334	1 DPOB RAT	P06766 rattus norv
26	21	100.0	342	1 AOP7_HUMAN	O14520 homo sapien
27	21	100.0	415	1 MUAL_THETN	Q8rd88 thermoanaer
28	21	100.0	430	1 RFEX_SALTY	P26400 salmonella
29	21	100.0	474	1 SAHH_RALSO	Q8y387 ralstonia s
30	21	100.0	502	1 TBG CANAL	O93807 candida alb
31	21	100.0	528	1 CTKI_YEAST	Q03957 saccharomyc
32	21	100.0	562	1 SYR_LACPL	Q88x53 lactobacill
33	21	100.0	563	1 SYR_ENTFA	Q831n1 enterococcu

ALIGNMENTS

RESULT 1				
MINE AGRT5	STANDARD;	PRT;	86 AA.	
AC Q8UAX0;				
DT 10-OCT-2003 (Rel. 42, Last sequence update)				
DT 10-OCT-2003 (Rel. 42, Last annotation update)				
DE Cell division topological specificity factor.				
GN MINE OR ATU3247 OR AGR_L3134.				
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).				
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.				
OX NCBI_TaxID=176299;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=21608550; PubMed=11743193;				
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,				
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,				
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,				
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,				
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,				
RA Raymond C., Rouse G., Saenbhimachak C., Wu Z., Romero P., Gordon D.,				
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,				
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,				
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,				
RA Nester E.W.;				
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens				
RT C58.";				
RT Science 294:2317-2323 (2001).				
RN [2]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=21608551; PubMed=11743194;				
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,				
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,				
RA Houniel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,				
RA Wollam C., Allington M., Doughty D., Scott C., Lappas C., Markelz B.,				
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,				
RA Cielo C., Slater S.;				
RT "Genome sequence of the plant pathogen and biotechnology agent				
RT Agrobacterium tumefaciens C58.";				
RT Science 294:2323-2328(2001).				
CC -!- FUNCTION: Prevents the cell division inhibition by proteins minC				
CC and mind at internal division sites while permitting inhibition at				
CC polar sites. This ensures cell division at the proper site by				
CC restricting the formation of a division septum at the midpoint of				
CC the long axis of the cell (By similarity).				
CC -!- SIMILARITY: Belongs to the minE family.				
CC -----				
CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way				
CC modified and this statement is not removed. Usage by and for commercial				
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC or send an email to license@isb-sib.ch).				
CC -----				

Q9cel2 lactococcus
P38788 saccharomyc
O40545 nicotiana t
P22317 alcaligenes
Q11174 caenorhabdi
Q08759 xenopus lae
O76031 homo sapien
Q9jhs4 mus musculu
P06876 mus musculu
P46200 bos taurus
P10242 homo sapien
P01103 gallus galli

34 21 100.0 564 1 SYR_LACLA
35 21 100.0 572 1 YHM4_YEAST
36 21 100.0 593 1 KPYA_TOBAC
37 21 100.0 602 1 HOXF_ALCEU
38 21 100.0 617 1 CHIT_CABEL
39 21 100.0 624 1 MYB_XENLA
40 21 100.0 633 1 CLPX_MOUSE
41 21 100.0 634 1 CLPX_MOUSE
42 21 100.0 636 1 MYB_MOUSE
43 21 100.0 640 1 MYB_BOVIN
44 21 100.0 640 1 MYB_HUMAN
45 21 100.0 641 1 MYB_CHICK

```

DR EMBL; AE009254; AAL44063.1; -.
DR EMBL; AE008359; AAK90142.1; -.
DR PIR; A12955; A12955.
DR PIR; D98327; D98327.
DR HAMAP; MF 00262; -.
DR InterPro; IPR005527; Mine.
DR Pfam; PF03776; Mine; 1.
DR TIGRFAMs; TIGR01215; mine; 1.
KW Cell division; Complete proteome.
SQ SEQUENCE 86 AA; 9703 MW; B0E274F6A48D52F2 CRC64;

  Query Match      100.0%; Score 21; DB 1; Length 86;
  Best Local Similarity 100.0%; Pred. No. 40;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 45 SVIAK 49

RESULT 2
RL27_MYCPE
ID RL27_MYCPE STANDARD; PRT; 95 AA.
AC Q8EYV5;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 50S ribosomal protein L27.
GN RPLA OR MYPE4440.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes, Mollicutes, Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22354719; PubMed=1246555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans."
RL Nucleic Acids Res. 30:5293-5300(2002).
CC -!- SIMILARITY: Belongs to the L27P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP004171; BAC44234.1; -.
DR HAMAP; MF 00539; -.
DR InterPro; IPR001684; Ribosomal L27.
DR Pfam; PF01016; Ribosomal L27; 1.
DR PRINTS; PR00063; RIBOSOMAL_L27.
DR ProDom; PD003114; Ribosomal_L27; 1.
DR TIGRFAMs; TIGR00062; L27; 1.
DR PROSITE; PS00831; RIBOSOMAL_L27; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 95 AA; 10464 MW; 735D951C94B7A730 CRC64;

  Query Match      100.0%; Score 21; DB 1; Length 95;
  Best Local Similarity 100.0%; Pred. No. 44;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 91 SVIAK 95

RESULT 3
SR19_METAC

```

```

ID SR19_METAC STANDARD; PRT; 101 AA.
AC Q8TTV3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Signal recognition particle 19 kDa protein (SRP19).
GN SRP19 OR MA0292.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=1193238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
CC -!- FUNCTION: Signal-recognition-particle assembly, binds directly to
CC 7S RNA and mediates binding of the 54 kDa subunit of the SRP (By
CC similarity).
CC -!- SUBUNIT: Archaeal signal recognition particle consists of a 7S RNA
CC molecule of 300 nucleotides and two protein subunits: SRP54 and
CC SRP19 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the SRP19 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010688; AAM03745.1; ALT_INIT.
DR HAMAP; MF_00305; -.
DR InterPro; IPR002778; SRP19.
DR Pfam; PF01922; SRP19; 1.
DR ProDom; PD006609; SRP19; 1.
DR Signal recognition particle; RNA-binding; Ribonucleoprotein;
KW Complete proteome.
SQ SEQUENCE 101 AA; 11415 MW; 8DA2E31AAA9594C3 CRC64;

  Query Match      100.0%; Score 21; DB 1; Length 101;
  Best Local Similarity 100.0%; Pred. No. 46;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 79 SVIAK 83

RESULT 4
SR19_METMA
ID SR19_METMA STANDARD; PRT; 101 AA.
AC Q8PFM7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Signal recognition particle 19 kDa protein (SRP19).
GN SRP19 OR MM1557.

```

```

OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / Goli / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierler A., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
CC -!- FUNCTION: Signal-recognition-particle assembly, binds directly to
CC 7S RNA and mediates binding of the 54 kDa subunit of the SRP (By
CC similarity).
CC -!- SUBUNIT: Archaeal signal recognition particle consists of a 7S RNA
CC molecule of 300 nucleotides and two protein subunits: SRP54 and
CC SRP19 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the SRP19 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE013390; AAM31253.1; -.
DR HAMAP; MF_00305; -.
DR InterPro; IPR002778; SRP19.
DR Pfam; PF01922; SRP19; 1.
DR ProDom; PD006609; SRP19; 1.
KW Signal recognition particle; RNA-binding; Ribonucleoprotein;
KW Complete proteome.
SQ SEQUENCE 101 AA; 11377 MW; 3F9235C41CF68C74 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
DB 79 SVIAK 83

RESULT 5
YA7B HAEIN STANDARD; PRT; 134 AA.
AC Q57425; P96338;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H11077.1.
GN H11077.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

```

```

RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Eritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.J., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512 (1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: WEAK, TO BACTERIAL PNUC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32788; AAC22744.1; -.
DR TIGR; H11077.1; -.
DR InterPro; IPR006419; NMN trans PnuC.
DR Pfam; PF04973; NMN_trans_PnuC; 1.
DR TIGRFAMs; TIGR01528; NMN_trans_PnuC; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 23 43
FT TRANSMEM 81 101
FT TRANSMEM 113 133
FT TRANSMEM 134 AA; 14415 MW; 70C1620F8D0E5BF CRC64;
SQ SEQUENCE 134 AA; 14415 MW; 70C1620F8D0E5BF CRC64;

Query Match 100.0%; Score 21; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
DB 75 SVIAK 79

RESULT 6
PUR3 STAAH STANDARD; PRT; 188 AA.
AC Q99V25;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR
DE transformylase) (5'-phosphoribosylglycinamide transformylase).
GN PURN OR SAVI072 OR SA0924.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MU50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240 (2001).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + N(1)-(5-phospho-D-
CC ribosyl)glycinamide = tetrahydrofolate + N(2)-formyl-N(1)-(5-
CC phospho-D-ribosyl)glycinamide
CC -!- PATHWAY: De novo purine biosynthesis; third step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: TO OTHER GART FROM BACTERIA AND EUKARYOTES.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AP003361; BAB57234.1; -;
 DR EMBL; AP003132; BAB42169.1; -;
 DR PIR; F89876; F89876.
 DR HSR; F08179; IGAR.
 DR SWISS-2DPAGE; Q99V25; STAA.
 DR InterPro; IPR002376; formyl transf.
 DR Pfam; PF00551; formyl transf; 1.
 DR PROSITE; PS00373; GART; FALSE_NEG.
 KW Purine biosynthesis; Transferase; Complete proteome.
 FT ACT_SITE 146 BY SIMILARITY.
 FT ACT_SITE 146
 SQ SEQUENCE 188 AA; 21166 MW; F0364618F275FA30 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 188;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 Db 181 SVIAK 185

RESULT 7

ID PUR3_STAAW STANDARD; PRT; 188 AA.
 AC Q8NX89;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR
 DE DE transformylase) (5'-phosphoribosylglycinamide transformylase).
 GN PURN OR M00955.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 [1]

SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
 RA Yanamoto K., Hiramatsu K.;
 RT "Genome and virulence determinants of high virulence community-
 RT acquired MRSA.";
 RL Lancet 359:1819-1827(2002).
 CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + N(1)-(5-phospho-D-
 CC ribosyl)glycinamide = tetrahydrofolate + N(2)-formyl-N(1)-(5-
 CC phospho-D-ribosyl)glycinamide.
 CC -!- PATHWAY: De novo purine biosynthesis; third step.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SIMILARITY: TO OTHER GART FROM BACTERIA AND EUKARYOTES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AP004825; BAB94820.1; -;
 DR InterPro; IPR002376; formyl transf.
 DR Pfam; PF00551; formyl transf; 1.
 DR PROSITE; PS00373; GART; FALSE_NEG.
 KW Purine biosynthesis; Transferase; Complete proteome.
 FT ACT_SITE 146 BY SIMILARITY.
 FT ACT_SITE 146
 SQ SEQUENCE 188 AA; 21153 MW; D034134258D89AEE CRC64;

Query Match 100.0%; Score 21; DB 1; Length 188;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 Db 181 SVIAK 185

RESULT 8

ID RNH2_SYNPX STANDARD; PRT; 224 AA.
 AC Q7U4C6;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ribonuclease HII (EC 3.1.26.4) (RNase HII).
 DE RNHB OR SYNW2144.
 GN Synechococcus sp. (strain WH8102).
 OS Synechococcus sp. (strain WH8102).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=84588;
 RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=22825697; PubMed=12917641;
 RA Palenik B., Brahamsha B., Larimer F.W., Land M., Hauser L., Chain P.,
 RA Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,
 RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;
 RT "The genome of a motile marine Synechococcus.";
 RL Nature 424:1037-1042(2003).
 CC -!- FUNCTION: This enzyme is an endonuclease that degrades the RNA of
 CC RNA-DNA hybrids specifically (By similarity).
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
 CC phosphomonocester.
 CC -!- COFACTOR: Manganese (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -!- SIMILARITY: Belongs to the RNase HII family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; BX569694; CAE08659.1; ALT_INIT.
 DR HAMAP; MF_00052; -; 1.
 DR InterPro; IPR001352; RNase HII/HIII.
 DR Pfam; PF01351; RNase HII; 1.
 KW Hydrolase; Nuclease; Endonuclease; Manganese; Complete proteome.
 FT ACT_SITE 42 42 BY SIMILARITY.
 FT ACT_SITE 138 138 BY SIMILARITY.
 FT ACT_SITE 157 157 BY SIMILARITY.
 SQ SEQUENCE 224 AA; 24310 MW; 7F21360ABC4C54AB CRC64;

Query Match 100.0%; Score 21; DB 1; Length 224;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
 |||||
 Db 167 SVIAK 171

RESULT 9

ID VGLL_HSV2H STANDARD; PRT; 224 AA.
 AC P28278;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Glycoprotein L precursor.

```

GN GL OR UL1.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2.";
RL J. Gen. Virol. 72:3057-3075 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ASSOCIATED WITH GLYCOPROTEIN H (GH) TO FORM A COMPLEX
CC -!- IMPORTANT FOR INFECTION AND CELL FUSION (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to the herpesviruses glycoprotein L family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10470; BAA01264.1; -.
DR EMBL; Z86099; CAB06761.1; -.
DR PIR; J01494; WVBEGH.
DR InterPro; IPR007923; Herpes_UL1.
DR Pfam; PF05259; Herpes_UL1; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 224 GLYCOPROTEIN L.
FT CAROHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 224 AA; 25192 MW; CDS85849250D7C1F CRC64;
SQ
Query Match 100.0%; Score 21; DB 1; Length 224;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVIAK 5
DB 30 SVIAK 34
RESULT 10
RS3_THEAC
ID RS3_THEAC STANDARD; PRT; 225 AA.
AC Q9HIR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3P.
GN RPS3P OR TAI265.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513 (2000).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head (By
CC similarity).
CC

```

```

CC -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445067; CAC12389.1; -.
DR HAMAP; MF 01309; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR009019; KH_prok.
DR InterPro; IPR004044; KH_TYPE_2.
DR InterPro; IPR001351; Ribosomal_S3_C.
DR InterPro; IPR005703; S3_euk_arch.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF00189; Ribosomal_S3_C; 1.
DR SMART; SM00322; KH; 1.
DR TIGRFAMs; TIGR01008; rpsC_E_A; 1.
DR PROSITE; PS0823; KH_TYPE_2; 1.
DR PROSITE; PS00548; RIBOSOMAL_S3; FALSE NEG.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 16 85 KH TYPE-2.
FT SEQUENCE 225 AA; 24726 MW; FE2B220091017F CRC64;
SQ
Query Match 100.0%; Score 21; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVIAK 5
DB 92 SVIAK 96
RESULT 11
RS3_THEVO
ID RS3_THEVO STANDARD; PRT; 225 AA.
AC Q97BX1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3P.
GN RPS3P OR TV0334 OR TVG0336522.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262 (2000).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC

```



```

ID VNST_PTPV STANDARD; PRT; 250 AA.
AC P03516;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nonstructural protein NS-S.
OS Punta toro phlebovirus.
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus.
OX NCBI_TaxID=11587;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84276006; PubMed=6087547;
RA Ihara T., Akashi H., Bishop D.H.L.;
RT "Novel coding strategy (ambisense genomic RNA) revealed by sequence
RL analyses of Punta toro phlebovirus S RNA.";
RL Virology 136:293-306(1984).
CC -!- MISCELLANEOUS: This protein may be a transcriptase component.
CC -!- SIMILARITY: NS-S FROM PUNTA TORO, RIFT VALLEY FEVER, SANDFLY FEVER
CC SICILIAN, TOSCANA, AND UKUNIEMI VIRUSES ARE EVOLUTIONARY RELATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02736; AAA47115.1; -.
DR PIR; A04108; MNVUPT.
KW Nonstructural protein; Transcription.
SQ SEQUENCE 250 AA; 29097 MW; 2C8909A1EDAD90D7 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
DB 55 SVIAK 59

RESULT 15
EXBB_XANCP STANDARD; PRT; 253 AA.
AC O34260;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biopolymer transport exbB protein.
GN EXBB OR XCC0009.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B100;
RX MEDLINE=98037510; PubMed=9371459;
RA Wiggerich H.G., Klaue B., Koeplin R., Priefer U.B., Puehler A.;
RT "Unusual structure of the tonB-exb DNA region of Xanthomonas
RT campestris pv. campestris: tonB, exbB, and exbD are essential for
RT ferric iron uptake, but exbD is not.";
RL J. Bacteriol. 179:7103-7110(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33913 / NCPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camnava F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,

```

```

RA Paria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- FUNCTION: Involved in the tonB-dependent energy-dependent
CC transport of various receptor-bound substrates. Protects exbD from
CC proteolytic degradation and functionally stabilizes tonB (By
CC similarity).
CC -!- SUBUNIT: The accessory proteins exbB and exbD seem to form a
CC complex with tonB (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -!- SIMILARITY: Belongs to the exbB / tonQ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95386; CAB08609.1; -.
DR EMBL; AE012094; AAM39328.1; -.
DR InterPro; IPR002898; Mota_ExbB.
DR Pfam; PF01618; Mota_ExbB; 1.
KW Transport; Protein transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 39
FT TRANSMEM 59
FT TRANSMEM 163
FT TRANSMEM 183
FT TRANSMEM 204
FT TRANSMEM 224
FT TRANSMEM 244
FT TRANSMEM 264
FT TRANSMEM 284
FT TRANSMEM 304
FT TRANSMEM 324
FT TRANSMEM 344
FT TRANSMEM 364
FT TRANSMEM 384
FT TRANSMEM 404
FT TRANSMEM 424
FT TRANSMEM 444
FT TRANSMEM 464
FT TRANSMEM 484
FT TRANSMEM 504
FT TRANSMEM 524
FT TRANSMEM 544
FT TRANSMEM 564
FT TRANSMEM 584
FT TRANSMEM 604
FT TRANSMEM 624
FT TRANSMEM 644
FT TRANSMEM 664
FT TRANSMEM 684
FT TRANSMEM 704
FT TRANSMEM 724
FT TRANSMEM 744
FT TRANSMEM 764
FT TRANSMEM 784
FT TRANSMEM 804
FT TRANSMEM 824
FT TRANSMEM 844
FT TRANSMEM 864
FT TRANSMEM 884
FT TRANSMEM 904
FT TRANSMEM 924
FT TRANSMEM 944
FT TRANSMEM 964
FT TRANSMEM 984
FT TRANSMEM 1004
FT TRANSMEM 1024
FT TRANSMEM 1044
FT TRANSMEM 1064
FT TRANSMEM 1084
FT TRANSMEM 1104
FT TRANSMEM 1124
FT TRANSMEM 1144
FT TRANSMEM 1164
FT TRANSMEM 1184
FT TRANSMEM 1204
FT TRANSMEM 1224
FT TRANSMEM 1244
FT TRANSMEM 1264
FT TRANSMEM 1284
FT TRANSMEM 1304
FT TRANSMEM 1324
FT TRANSMEM 1344
FT TRANSMEM 1364
FT TRANSMEM 1384
FT TRANSMEM 1404
FT TRANSMEM 1424
FT TRANSMEM 1444
FT TRANSMEM 1464
FT TRANSMEM 1484
FT TRANSMEM 1504
FT TRANSMEM 1524
FT TRANSMEM 1544
FT TRANSMEM 1564
FT TRANSMEM 1584
FT TRANSMEM 1604
FT TRANSMEM 1624
FT TRANSMEM 1644
FT TRANSMEM 1664
FT TRANSMEM 1684
FT TRANSMEM 1704
FT TRANSMEM 1724
FT TRANSMEM 1744
FT TRANSMEM 1764
FT TRANSMEM 1784
FT TRANSMEM 1804
FT TRANSMEM 1824
FT TRANSMEM 1844
FT TRANSMEM 1864
FT TRANSMEM 1884
FT TRANSMEM 1904
FT TRANSMEM 1924
FT TRANSMEM 1944
FT TRANSMEM 1964
FT TRANSMEM 1984
FT TRANSMEM 2004
FT TRANSMEM 2024
FT TRANSMEM 2044
FT TRANSMEM 2064
FT TRANSMEM 2084
FT TRANSMEM 2104
FT TRANSMEM 2124
FT TRANSMEM 2144
FT TRANSMEM 2164
FT TRANSMEM 2184
FT TRANSMEM 2204
FT TRANSMEM 2224
FT TRANSMEM 2244
FT TRANSMEM 2264
FT TRANSMEM 2284
FT TRANSMEM 2304
FT TRANSMEM 2324
FT TRANSMEM 2344
FT TRANSMEM 2364
FT TRANSMEM 2384
FT TRANSMEM 2404
FT TRANSMEM 2424
FT TRANSMEM 2444
FT TRANSMEM 2464
FT TRANSMEM 2484
FT TRANSMEM 2504
FT TRANSMEM 2524
FT TRANSMEM 2544
FT TRANSMEM 2564
FT TRANSMEM 2584
FT TRANSMEM 2604
FT TRANSMEM 2624
FT TRANSMEM 2644
FT TRANSMEM 2664
FT TRANSMEM 2684
FT TRANSMEM 2704
FT TRANSMEM 2724
FT TRANSMEM 2744
FT TRANSMEM 2764
FT TRANSMEM 2784
FT TRANSMEM 2804
FT TRANSMEM 2824
FT TRANSMEM 2844
FT TRANSMEM 2864
FT TRANSMEM 2884
FT TRANSMEM 2904
FT TRANSMEM 2924
FT TRANSMEM 2944
FT TRANSMEM 2964
FT TRANSMEM 2984
FT TRANSMEM 3004
FT TRANSMEM 3024
FT TRANSMEM 3044
FT TRANSMEM 3064
FT TRANSMEM 3084
FT TRANSMEM 3104
FT TRANSMEM 3124
FT TRANSMEM 3144
FT TRANSMEM 3164
FT TRANSMEM 3184
FT TRANSMEM 3204
FT TRANSMEM 3224
FT TRANSMEM 3244
FT TRANSMEM 3264
FT TRANSMEM 3284
FT TRANSMEM 3304
FT TRANSMEM 3324
FT TRANSMEM 3344
FT TRANSMEM 3364
FT TRANSMEM 3384
FT TRANSMEM 3404
FT TRANSMEM 3424
FT TRANSMEM 3444
FT TRANSMEM 3464
FT TRANSMEM 3484
FT TRANSMEM 3504
FT TRANSMEM 3524
FT TRANSMEM 3544
FT TRANSMEM 3564
FT TRANSMEM 3584
FT TRANSMEM 3604
FT TRANSMEM 3624
FT TRANSMEM 3644
FT TRANSMEM 3664
FT TRANSMEM 3684
FT TRANSMEM 3704
FT TRANSMEM 3724
FT TRANSMEM 3744
FT TRANSMEM 3764
FT TRANSMEM 3784
FT TRANSMEM 3804
FT TRANSMEM 3824
FT TRANSMEM 3844
FT TRANSMEM 3864
FT TRANSMEM 3884
FT TRANSMEM 3904
FT TRANSMEM 3924
FT TRANSMEM 3944
FT TRANSMEM 3964
FT TRANSMEM 3984
FT TRANSMEM 4004
FT TRANSMEM 4024
FT TRANSMEM 4044
FT TRANSMEM 4064
FT TRANSMEM 4084
FT TRANSMEM 4104
FT TRANSMEM 4124
FT TRANSMEM 4144
FT TRANSMEM 4164
FT TRANSMEM 4184
FT TRANSMEM 4204
FT TRANSMEM 4224
FT TRANSMEM 4244
FT TRANSMEM 4264
FT TRANSMEM 4284
FT TRANSMEM 4304
FT TRANSMEM 4324
FT TRANSMEM 4344
FT TRANSMEM 4364
FT TRANSMEM 4384
FT TRANSMEM 4404
FT TRANSMEM 4424
FT TRANSMEM 4444
FT TRANSMEM 4464
FT TRANSMEM 4484
FT TRANSMEM 4504
FT TRANSMEM 4524
FT TRANSMEM 4544
FT TRANSMEM 4564
FT TRANSMEM 4584
FT TRANSMEM 4604
FT TRANSMEM 4624
FT TRANSMEM 4644
FT TRANSMEM 4664
FT TRANSMEM 4684
FT TRANSMEM 4704
FT TRANSMEM 4724
FT TRANSMEM 4744
FT TRANSMEM 4764
FT TRANSMEM 4784
FT TRANSMEM 4804
FT TRANSMEM 4824
FT TRANSMEM 4844
FT TRANSMEM 4864
FT TRANSMEM 4884
FT TRANSMEM 4904
FT TRANSMEM 4924
FT TRANSMEM 4944
FT TRANSMEM 4964
FT TRANSMEM 4984
FT TRANSMEM 5004
FT TRANSMEM 5024
FT TRANSMEM 5044
FT TRANSMEM 5064
FT TRANSMEM 5084
FT TRANSMEM 5104
FT TRANSMEM 5124
FT TRANSMEM 5144
FT TRANSMEM 5164
FT TRANSMEM 5184
FT TRANSMEM 5204
FT TRANSMEM 5224
FT TRANSMEM 5244
FT TRANSMEM 5264
FT TRANSMEM 5284
FT TRANSMEM 5304
FT TRANSMEM 5324
FT TRANSMEM 5344
FT TRANSMEM 5364
FT TRANSMEM 5384
FT TRANSMEM 5404
FT TRANSMEM 5424
FT TRANSMEM 5444
FT TRANSMEM 5464
FT TRANSMEM 5484
FT TRANSMEM 5504
FT TRANSMEM 5524
FT TRANSMEM 5544
FT TRANSMEM 5564
FT TRANSMEM 5584
FT TRANSMEM 5604
FT TRANSMEM 5624
FT TRANSMEM 5644
FT TRANSMEM 5664
FT TRANSMEM 5684
FT TRANSMEM 5704
FT TRANSMEM 5724
FT TRANSMEM 5744
FT TRANSMEM 5764
FT TRANSMEM 5784
FT TRANSMEM 5804
FT TRANSMEM 5824
FT TRANSMEM 5844
FT TRANSMEM 5864
FT TRANSMEM 5884
FT TRANSMEM 5904
FT TRANSMEM 5924
FT TRANSMEM 5944
FT TRANSMEM 5964
FT TRANSMEM 5984
FT TRANSMEM 6004
FT TRANSMEM 6024
FT TRANSMEM 6044
FT TRANSMEM 6064
FT TRANSMEM 6084
FT TRANSMEM 6104
FT TRANSMEM 6124
FT TRANSMEM 6144
FT TRANSMEM 6164
FT TRANSMEM 6184
FT TRANSMEM 6204
FT TRANSMEM 6224
FT TRANSMEM 6244
FT TRANSMEM 6264
FT TRANSMEM 6284
FT TRANSMEM 6304
FT TRANSMEM 6324
FT TRANSMEM 6344
FT TRANSMEM 6364
FT TRANSMEM 6384
FT TRANSMEM 6404
FT TRANSMEM 6424
FT TRANSMEM 6444
FT TRANSMEM 6464
FT TRANSMEM 6484
FT TRANSMEM 6504
FT TRANSMEM 6524
FT TRANSMEM 6544
FT TRANSMEM 6564
FT TRANSMEM 6584
FT TRANSMEM 6604
FT TRANSMEM 6624
FT TRANSMEM 6644
FT TRANSMEM 6664
FT TRANSMEM 6684
FT TRANSMEM 6704
FT TRANSMEM 6724
FT TRANSMEM 6744
FT TRANSMEM 6764
FT TRANSMEM 6784
FT TRANSMEM 6804
FT TRANSMEM 6824
FT TRANSMEM 6844
FT TRANSMEM 6864
FT TRANSMEM 6884
FT TRANSMEM 6904
FT TRANSMEM 6924
FT TRANSMEM 6944
FT TRANSMEM 6964
FT TRANSMEM 6984
FT TRANSMEM 7004
FT TRANSMEM 7024
FT TRANSMEM 7044
FT TRANSMEM 7064
FT TRANSMEM 7084
FT TRANSMEM 7104
FT TRANSMEM 7124
FT TRANSMEM 7144
FT TRANSMEM 7164
FT TRANSMEM 7184
FT TRANSMEM 7204
FT TRANSMEM 7224
FT TRANSMEM 7244
FT TRANSMEM 7264
FT TRANSMEM 7284
FT TRANSMEM 7304
FT TRANSMEM 7324
FT TRANSMEM 7344
FT TRANSMEM 7364
FT TRANSMEM 7384
FT TRANSMEM 7404
FT TRANSMEM 7424
FT TRANSMEM 7444
FT TRANSMEM 7464
FT TRANSMEM 7484
FT TRANSMEM 7504
FT TRANSMEM 7524
FT TRANSMEM 7544
FT TRANSMEM 7564
FT TRANSMEM 7584
FT TRANSMEM 7604
FT TRANSMEM 7624
FT TRANSMEM 7644
FT TRANSMEM 7664
FT TRANSMEM 7684
FT TRANSMEM 7704
FT TRANSMEM 7724
FT TRANSMEM 7744
FT TRANSMEM 7764
FT TRANSMEM 7784
FT TRANSMEM 7804
FT TRANSMEM 7824
FT TRANSMEM 7844
FT TRANSMEM 7864
FT TRANSMEM 7884
FT TRANSMEM 7904
FT TRANSMEM 7924
FT TRANSMEM 7944
FT TRANSMEM 7964
FT TRANSMEM 7984
FT TRANSMEM 8004
FT TRANSMEM 8024
FT TRANSMEM 8044
FT TRANSMEM 8064
FT TRANSMEM 8084
FT TRANSMEM 8104
FT TRANSMEM 8124
FT TRANSMEM 8144
FT TRANSMEM 8164
FT TRANSMEM 8184
FT TRANSMEM 8204
FT TRANSMEM 8224
FT TRANSMEM 8244
FT TRANSMEM 8264
FT TRANSMEM 8284
FT TRANSMEM 8304
FT TRANSMEM 8324
FT TRANSMEM 8344
FT TRANSMEM 8364
FT TRANSMEM 8384
FT TRANSMEM 8404
FT TRANSMEM 8424
FT TRANSMEM 8444
FT TRANSMEM 8464
FT TRANSMEM 8484
FT TRANSMEM 8504
FT TRANSMEM 8524
FT TRANSMEM 8544
FT TRANSMEM 8564
FT TRANSMEM 8584
FT TRANSMEM 8604
FT TRANSMEM 8624
FT TRANSMEM 8644
FT TRANSMEM 8664
FT TRANSMEM 8684
FT TRANSMEM 8704
FT TRANSMEM 8724
FT TRANSMEM 8744
FT TRANSMEM 8764
FT TRANSMEM 8784
FT TRANSMEM 8804
FT TRANSMEM 8824
FT TRANSMEM 8844
FT TRANSMEM 8864
FT TRANSMEM 8884
FT TRANSMEM 8904
FT TRANSMEM 8924
FT TRANSMEM 8944
FT TRANSMEM 8964
FT TRANSMEM 8984
FT TRANSMEM 9004
FT TRANSMEM 9024
FT TRANSMEM 9044
FT TRANSMEM 9064
FT TRANSMEM 9084
FT TRANSMEM 9104
FT TRANSMEM 9124
FT TRANSMEM 9144
FT TRANSMEM 9164
FT TRANSMEM 9184
FT TRANSMEM 9204
FT TRANSMEM 9224
FT TRANSMEM 9244
FT TRANSMEM 9264
FT TRANSMEM 9284
FT TRANSMEM 9304
FT TRANSMEM 9324
FT TRANSMEM 9344
FT TRANSMEM 9364
FT TRANSMEM 9384
FT TRANSMEM 9404
FT TRANSMEM 9424
FT TRANSMEM 9444
FT TRANSMEM 9464
FT TRANSMEM 9484
FT TRANSMEM 9504
FT TRANSMEM 9524
FT TRANSMEM 9544
FT TRANSMEM 9564
FT TRANSMEM 9584
FT TRANSMEM 9604
FT TRANSMEM 9624
FT TRANSMEM 9644
FT TRANSMEM 9664
FT TRANSMEM 9684
FT TRANSMEM 9704
FT TRANSMEM 9724
FT TRANSMEM 9744
FT TRANSMEM 9764
FT TRANSMEM 9784
FT TRANSMEM 9804
FT TRANSMEM 9824
FT TRANSMEM 9844
FT TRANSMEM 9864
FT TRANSMEM 9884
FT TRANSMEM 9904
FT TRANSMEM 9924
FT TRANSMEM 9944
FT TRANSMEM 9964
FT TRANSMEM 9984
FT TRANSMEM 10004
FT TRANSMEM 10024
FT TRANSMEM 10044
FT TRANSMEM 10064
FT TRANSMEM 10084
FT TRANSMEM 10104
FT TRANSMEM 10124
FT TRANSMEM 10144
FT TRANSMEM 10164
FT TRANSMEM 10184
FT TRANSMEM 10204
FT TRANSMEM 10224
FT TRANSMEM 10244
FT TRANSMEM 10264
FT TRANSMEM 10284
FT TRANSMEM 10304
FT TRANSMEM 10324
FT TRANSMEM 10344
FT TRANSMEM 10364
FT TRANSMEM 10384
FT TRANSMEM 10404
FT TRANSMEM 10424
FT TRANSMEM 10444
FT TRANSMEM 10464
FT TRANSMEM 10484
FT TRANSMEM 10504
FT TRANSMEM 10524
FT TRANSMEM 10544
FT TRANSMEM 10564
FT TRANSMEM 10584
FT TRANSMEM 10604
FT TRANSMEM 10624
FT TRANSMEM 10644
FT TRANSMEM 10664
FT TRANSMEM 10684
FT TRANSMEM 10704
FT TRANSMEM 10724
FT TRANSMEM 10744
FT TRANSMEM 10764
FT TRANSMEM 10784
FT TRANSMEM 10804
FT TRANSMEM 10824
FT TRANSMEM 10844
FT TRANSMEM 10864
FT TRANSMEM 10884
FT TRANSMEM 10904
FT TRANSMEM 10924
FT TRANSMEM 10944
FT TRANSMEM 10964
FT TRANSMEM 10984
FT TRANSMEM 11004
FT TRANSMEM 11024
FT TRANSMEM 11044
FT TRANSMEM 11064
FT TRANSMEM 11084
FT TRANSMEM 11104
FT TRANSMEM 11124
FT TRANSMEM 11144
FT TRANSMEM 11164
FT TRANSMEM 11184
FT TRANSMEM 11204
FT TRANSMEM 11224
FT TRANSMEM 11244
FT TRANSMEM 11264
FT TRANSMEM 11284
FT TRANSMEM 11304
FT TRANSMEM 11324
FT TRANSMEM 11344
FT TRANSMEM 11364
FT TRANSMEM 11384
FT TRANSMEM 11404
FT TRANSMEM 11424
FT TRANSMEM 11444
FT TRANSMEM 11464
FT TRANSMEM 11484
FT TRANSMEM 11504
FT TRANSMEM 11524
FT TRANSMEM 11544
FT TRANSMEM 11564
FT TRANSMEM 11584
FT TRANSMEM 11604
FT TRANSMEM 11624
FT TRANSMEM 11644
FT TRANSMEM 11664
FT TRANSMEM 11684
FT TRANSMEM 11704
FT TRANSMEM 11724
FT TRANSMEM 11744
FT TRANSMEM 11764
FT TRANSMEM 11784
FT TRANSMEM 11804
FT TRANSMEM 11824
FT TRANSMEM 11844
FT TRANSMEM 11864
FT TRANSMEM 11884
FT TRANSMEM 11904
FT TRANSMEM 11924
FT TRANSMEM 11944
FT TRANSMEM 11964
FT TRANSMEM 11984
FT TRANSMEM 12004
FT TRANSMEM 12024
FT TRANSMEM 12044
FT TRANSMEM 12064
FT TRANSMEM 12084
FT TRANSMEM 12104
FT TRANSMEM 12124
FT TRANSMEM 12144
FT TRANSMEM 12164
FT TRANSMEM 12184
FT TRANSMEM 12204
FT TRANSMEM 12224
FT TRANSMEM 12244
FT TRANSMEM 12264
FT TRANSMEM 12284
FT TRANSMEM 12304
FT TRANSMEM 12324
FT TRANSMEM 12344
FT TRANSMEM 12364
FT TRANSMEM 12384
FT TRANSMEM 12404
FT TRANSMEM 12424
FT TRANSMEM 12444
FT TRANSMEM 12464
FT TRANSMEM 12484
FT TRANSMEM 12504
FT TRANSMEM 12524
FT TRANSMEM 12544
FT TRANSMEM 12564
FT TRANSMEM 12584
FT TRANSMEM 12604
FT TRANSMEM 12624
FT TRANSMEM 12644
FT TRANSMEM 12664
FT TRANSMEM 12684
FT TRANSMEM 12704
FT TRANSMEM 12724
FT TRANSMEM 12744
FT TRANSMEM 12764
FT TRANSMEM 12784
FT TRANSMEM 12804
FT TRANSMEM 12824
FT TRANSMEM 12844
FT TRANSMEM 12864
FT TRANSMEM 12884
FT TRANSMEM 12904
FT TRANSMEM 12924
FT TRANSMEM 12944
FT TRANSMEM 12964
FT TRANSMEM 12984
FT TRANSMEM 13004
FT TRANSMEM 13024
FT TRANSMEM 13044
FT TRANSMEM 13064
FT TRANSMEM 13084
FT TRANSMEM 13104
FT TRANSMEM 13124
FT TRANSMEM 13144
FT TRANSMEM 13164
FT TRANSMEM 13184
FT TRANSMEM 13204
FT TRANSMEM 13224
FT TRANSMEM 13244
FT TRANSMEM 13264
FT TRANSMEM 13284
FT TRANSMEM 13304
FT TRANSMEM 13324
FT TRANSMEM 13344
FT TRANSMEM 13364
FT TRANSMEM 13384
FT TRANSMEM 13404
FT TRANSMEM 13424
FT TRANSMEM 13444
FT TRANSMEM 13464
FT TRANSMEM 13484
FT TRANSMEM 13504
FT TRANSMEM 13524
FT TRANSMEM 13544
FT TRANSMEM 13564
FT TRANSMEM 13584
FT TRANSMEM 13604
FT TRANSMEM 13624
FT TRANSMEM 13644
FT TRANSMEM 13664
FT TRANSMEM 13684
FT TRANSMEM 13704
FT TRANSMEM 13724
FT TRANSMEM 13744
FT TRANSMEM 13764
FT TRANSMEM 13784
FT TRANSMEM 13804
FT TRANSMEM 13824
FT TRANSMEM 13844
FT TRANSMEM 13864
FT TRANSMEM 13884
FT TRANSMEM 13904
FT TRANSMEM 13924
FT TRANSMEM 13944
FT TRANSMEM 13964
FT TRANSMEM 13984
FT TRANSMEM 14004
FT TRANSMEM 14024
FT TRANSMEM 14044
FT TRANSMEM 14064
FT TRANSMEM 14084
FT TRANSMEM 14104
FT TRANSMEM 14124
FT TRANSMEM 14144
FT TRANSMEM 14164
FT TRANSMEM 14184
FT TRANSMEM 14204
FT TRANSMEM 14224
FT TRANSMEM 14244
FT TRANSMEM 14264
FT TRANSMEM 14284
FT TRANSMEM 14304
FT TRANSMEM 14324
FT TRANSMEM 14344
FT TRANSMEM 14364
FT TRANSMEM 14384
FT TRANSMEM 14404
FT TRANSMEM 14424
FT TRANSMEM 14444
FT TRANSMEM 14464
FT TRANSMEM 14484
FT TRANSMEM 14504
FT TRANSMEM 14524
FT TRANSMEM 14544
FT TRANSMEM 14564
FT TRANSMEM 14584
FT TRANSMEM 14604
FT TRANSMEM 14624
FT TRANSMEM 14644
FT TRANSMEM 14664
FT TRANSMEM 14684
FT TRANSMEM 14704
FT TRANSMEM 14724
FT TRANSMEM 14744
FT TRANSMEM 14764
FT TRANSMEM 14784
FT TRANSMEM 14804
FT TRANSMEM 14824
FT TRANSMEM 14844
FT TRANSMEM 14864
FT TRANSMEM 14884
FT TRANSMEM 14904
FT TRANSMEM 14924
FT TRANSMEM 14944
FT TRANSMEM 14964
FT TRANSMEM 14984
FT TRANSMEM 15004
FT TRANSMEM 15024
FT TRANSMEM 15044
FT TRANSMEM 15064
FT TRANSMEM 15084
FT TRANSMEM 15104
FT TRANSMEM 15124
FT TRANSMEM 15144
FT TRANSMEM 15164
FT TRANSMEM 15184
FT TRANSMEM 15204
FT TRANSMEM 15224
FT TRANSMEM 15244
FT TRANSMEM 15264
FT TRANSMEM 15284
FT TRANSMEM 15304
FT TRANSMEM 15324
FT TRANSMEM 15344
FT TRANSMEM 15364
FT TRANSMEM 15384
FT TRANSMEM 15404
FT TRANSMEM 15424
FT TRANSMEM 15444
FT TRANSMEM 15464
FT TRANSMEM 15484
FT TRANSMEM 15504
FT TRANSMEM 15524
FT TRANSMEM 15544
FT TRANSMEM 15564
FT TRANSMEM 15584
FT TRANSMEM 15604
FT TRANSMEM 15624
FT TRANSMEM 15644
FT TRANSMEM 15664
FT TRANSMEM 15684
FT TRANSMEM 15704
FT TRANSMEM 15724
FT TRANSMEM 15744
FT TRANSMEM 15764
FT TRANSMEM 15784
FT TRANSMEM 15804
FT TRANSMEM 15824
FT TRANSMEM 15844
FT TRANSMEM 15864
FT TRANSMEM 15884
FT TRANSMEM 15904
FT TRANSMEM 15924
FT TRANSMEM 15944
FT TRANSMEM 15964
FT TRANSMEM 15984
FT TRANSMEM 16004
FT TRANSMEM 16024
FT TRANSMEM 16044
FT TRANSMEM 16064
FT TRANSMEM 16084
FT TRANSMEM 16104
FT TRANSMEM 16124
FT TRANSMEM 16144
FT TRANSMEM 16164
FT TRANSMEM 16184
FT TRANSMEM 16204
FT TRANSMEM 16224
FT TRANSMEM 16244
FT TRANSMEM 16264
FT TRANSMEM 16284
FT TRANSMEM 16304
FT TRANSMEM 16324
FT TRANSMEM 16344
FT TRANSMEM 16364
FT TRANSMEM 16384
FT TRANSMEM 16404
FT TRANSMEM 16424
FT TRANSMEM 16444
FT TRANSMEM 16464
FT TRANSMEM 16484
FT TRANSMEM 16504
FT TRANSMEM 16524
FT TRANSMEM 16544
FT TRANSMEM 16564
FT TRANSMEM 16584
FT TRANSMEM 16604
FT TRANSMEM 16624
FT TRANSMEM 16644
FT TRANSMEM 16664
FT TRANSMEM 16684
FT TRANSMEM 16704
FT TRANSMEM 16724
FT TRANSMEM 16744
FT TRANSMEM 16764
FT TRANSMEM 16784
FT TRANSMEM 16804
FT TRANSMEM 16824
FT TRANSMEM 16844
FT TRANSMEM 16864
FT TRANSMEM 16884
FT TRANSMEM 16904
FT TRANSMEM 16924
FT TRANSMEM 16944
FT TRANSMEM 16964
FT TRANSMEM 16984
FT TRANSMEM 17004
FT TRANSMEM 17024
FT TRANSMEM 17044
FT TRANSMEM 17064
FT TRANSMEM 17084
FT TRANSMEM 17104
FT TRANSMEM 17124
FT TRANSMEM 17144
FT TRANSMEM 17164
FT TRANSMEM 17184
FT TRANSMEM 17204
FT TRANSMEM 17224
FT TRANSMEM 17244
FT TRANSMEM 17264
FT TRANSMEM 17284
FT TRANSMEM 17304
FT TRANSMEM 17324
FT TRANSMEM 17344
FT TRANSMEM 17364
FT TRANSMEM 17384
FT TRANSMEM 17404
FT TRANSMEM 17424
FT TRANSMEM 17444
FT TRANSMEM 17464
FT TRANSMEM 17484
FT TRANSMEM 17504
FT TRANSMEM 17524
FT TRANSMEM 17544
FT TRANSMEM 17564
FT TRANSMEM 17584
FT TRANSMEM 17604
FT TRANSMEM 17624
FT TRANSMEM 17644
FT TRANSMEM 17664
FT TRANSMEM 17684
FT TRANSMEM 17704
FT TRANSMEM 17724
FT TRANSMEM 17744
FT TRANSMEM 17764
FT TRANSMEM 17784
FT TRANSMEM 17804
FT TRANSMEM 17824
FT TRANSMEM 17844
FT TRANSMEM 17864
FT TRANSMEM 17884
FT TRANSMEM 17904
FT TRANSMEM 17924
FT TRANSMEM 17944
FT TRANSMEM 17964
FT TRANSMEM 17984
FT TRANSMEM 18004
FT TRANSMEM 18024
FT TRANSMEM 18044
FT TRANSMEM 18064
FT TRANSMEM 18084
FT TRANSMEM 18104
FT TRANSMEM 18124
FT TRANSMEM 18144
FT TRANSMEM 18164
FT TRANSMEM 18184
FT TRANSMEM 18204
FT TRANSMEM 18224
FT TRANSMEM 18244
FT TRANSMEM 18264
FT TRANSMEM 18284
FT TRANSMEM 18304
FT TRANSMEM 18324
FT TRANSMEM 18344
FT TRANSMEM 18364
FT TRANSMEM 18384
FT TRANSMEM 18404
FT TRANSMEM 18424
FT TRANSMEM 18444
FT TRANSMEM 18464
FT TRANSMEM 18484
FT TRANSMEM 18504
FT TRANSMEM 18524
FT TRANSMEM 18544
FT TRANSMEM 18564
FT TRANSMEM 18584
FT TRANSMEM 18604
FT TRANSMEM 18624
FT TRANSMEM 18644
FT TRANSMEM 18664
FT TRANSMEM 18684
FT TRANSMEM 18704
FT TRANSMEM 18724
FT TRANSMEM 18744
FT TRANSMEM 18764
FT TRANSMEM 18784
FT TRANSMEM 18804
FT TRANSMEM 18824
FT TRANSMEM 18844
FT TRANSMEM 18864
FT TRANSMEM 18884
FT TRANSMEM 18904
FT TRANSMEM 18924
FT TRANSMEM 18944
FT TRANSMEM 18964
FT TRANSMEM 18984
FT TRANSMEM 19004
FT TRANSMEM 19024
FT TRANSMEM 19044
FT TRANSMEM 19064
FT TRANSMEM 19084
FT TRANSMEM 19104
FT TRANSMEM 19124
FT TRANSMEM 19144
FT TRANSMEM 19164
FT TRANSMEM 19184
FT TRANSMEM 19204
FT TRANSMEM 19224
FT TRANSMEM 19244
FT TRANSMEM 19264
FT TRANSMEM 19284
FT TRANSMEM 19304
FT TRANSMEM 19324
FT TRANSMEM 19344
FT TRANSMEM 19364
FT TRANSMEM 19384
FT TRANSMEM 19404
FT TRANSMEM 19424
FT TRANSMEM 19444
FT TRANSMEM 19464
FT TRANSMEM 19484
FT TRANSMEM 19504
FT TRANSMEM 19524
FT TRANSMEM 19544
FT TRANSMEM 19564
FT TRANSMEM 19584
FT TRANSMEM 19604
FT TRANSMEM 19624
FT TRANSMEM 19644
FT TRANSMEM 19664
FT TRANSMEM 19684
FT TRANSMEM 19704
FT TRANSMEM 19724
FT TRANSMEM 19744
FT TRANSMEM 19764
FT TRANSMEM 19784
FT TRANSMEM 19804
FT TRANSMEM 19824
FT TRANSMEM 19844
FT TRANSMEM 1986
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.30123 Seconds

(without alignments)
1212.385 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAX 5

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mbc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	63	Q9XT66	Q9xt66 canis famil
2	21	100.0	75	Q8CRL6	Q8crl6 staphylococ
3	21	100.0	95	Q8EVM5	Q8evw5 mycoplasma
4	21	100.0	98	Q8P4C6	Q8p4c6 xanthomonas
5	21	100.0	107	Q7YUB8	Q7yub8 aphelenchus
6	21	100.0	111	Q922Z7	Q922z7 mus musculu
7	21	100.0	126	Q8HAJ2	Q8haj2 bacterioph
8	21	100.0	126	Q8C8J3	Q8c8j3 mus musculu
9	21	100.0	132	Q17211	Q17211 caenorhabdi
10	21	100.0	135	Q8SC48	Q8sc48 stx2 conver
11	21	100.0	141	Q8E7B3	Q8e7b3 streptococc
12	21	100.0	141	Q89P48	Q89p48 bradyrhizob
13	21	100.0	152	Q8VK36	Q8vk36 mycobacteri
14	21	100.0	152	Q7U080	Q7u080 mycobacteri
15	21	100.0	161	Q9XJQ6	Q9xjq6 bacterioph
16	21	100.0	161	Q8ZVL5	Q8zvl5 pyrobaculum

17	21	100.0	162	9	Q8HA15	Q8ha15 bacterioph
18	21	100.0	164	16	Q8EXA9	Q8exa9 leptospira
19	21	100.0	171	16	Q8RE43	Q8re43 mycoplasma
20	21	100.0	172	9	Q8HAE9	Q8hae9 salmonella
21	21	100.0	175	3	Q00300	Q00300 ajellomyces
22	21	100.0	175	3	Q9P436	Q9p436 ajellomyces
23	21	100.0	175	3	Q9P433	Q9p433 ajellomyces
24	21	100.0	175	3	Q9P439	Q9p439 ajellomyces
25	21	100.0	175	3	Q9P435	Q9p435 ajellomyces
26	21	100.0	175	3	Q9P437	Q9p437 ajellomyces
27	21	100.0	175	3	Q9P434	Q9p434 ajellomyces
28	21	100.0	176	13	P87467	P87467 gallus gall
29	21	100.0	177	11	Q9CJW7	Q9cvj7 mus musculu
30	21	100.0	181	16	Q8YV51	Q8yv51 anabaena sp
31	21	100.0	188	16	Q9V25	Q9v25 staphylococ
32	21	100.0	188	16	Q8X89	Q8x89 staphylococ
33	21	100.0	191	11	P97753	P97753 mus sp. gag
34	21	100.0	192	16	Q8R852	Q8r852 thermococ
35	21	100.0	196	16	Q8G813	Q8g813 bifidobacte
36	21	100.0	197	16	Q9AAV8	Q9aav8 caulobacter
37	21	100.0	199	2	O52946	O52946 bacillus su
38	21	100.0	200	16	Q8Y9M4	Q8y9m4 listeria mo
39	21	100.0	203	10	Q9SW28	Q9sw28 arabidopsis
40	21	100.0	209	16	Q8FS11	Q8fs11 corynebacte
41	21	100.0	218	5	Q9N6J5	Q9n6j5 leishmania
42	21	100.0	219	16	Q9F8V9	Q9f8v9 agrobacteri
43	21	100.0	221	5	Q95P04	Q95p04 goniorpora t
44	21	100.0	223	2	Q9AHZ2	Q9ahz2 photorhabdu
45	21	100.0	224	12	Q69112	Q69112 herpes simp

ALIGNMENTS

RESULT 1	Q9XT66	PRELIMINARY;	PRT;	63 AA.
ID	Q9XT66;			
AC	Q9XT66;			
DT	01-NOV-1999 (TRENBLrel. 12, Created)			
DT	01-NOV-1999 (TRENBLrel. 12, Last sequence update)			
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE	C-myb (Fragment)			
GN	C-MYB.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99265967; PubMed=1031940;			
RA	Li R., Mignot E., Faraco J., Kadotani H., Cantanese J., Zhao B.,			
RA	Lin X., Hinton L., Ostrander E.A., Patterson D.F., de Jong P.J.;			
RT	"Construction and characterization of an eightfold redundant dog			
RT	genomic bacterial artificial chromosome library.";			
RL	Genomics 58:9-17(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Kodatani H., Mignot E.;			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
CC	-!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
CC	-!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.			
DR	EMBL; AF103748; AAD40574.1; -.			
DR	HSSP; P06876; 1MBG.			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0003677; F:DNA binding; IEA.			
DR	InterPro; IPR001005; Myb DNA binding.			
DR	Pfam; PF00249; myb DNA-binding; 2.			
DR	SMART; SM00717; SANT; 1.			
DR	PROSITE; PS00037; MYB_1; 1.			
DR	PROSITE; PS00334; MYB_2; 1.			
DR	PROSITE; PS00090; MYB_3; 1.			
KW	Nuclear protein.			
FT	NON_TER			1

```

FT NON_TER 63
SQ SEQUENCE 63 AA; 7707 MW; D8C96265802F3C9F CRC64;

Query Match 100.0%; Score 21; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 19 SVIAK 23

RESULT 2
Q8CRL6
ID Q8CRL6 PRELIMINARY; PRT; 75 AA.
AC Q8CRL6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
DE SEI742.
GN Staphylococcus epidermidis.
OS Bacteria; Firmicutes; Bacillales; Staphylococcus.
OC NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016749; AAO05341.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 75 AA; 8090 MW; 9B017C6D9C61D9F CRC64;

Query Match 100.0%; Score 21; DB 16; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 71 SVIAK 75

RESULT 3
Q8EVW5
ID Q8EVW5 PRELIMINARY; PRT; 95 AA.
AC Q8EVW5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ribosomal protein L27.
GN MYPE4440.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RA MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "the complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004171; BAC44234.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; P:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001684; Ribosomal_L27.
DR Pfam; PF01016; Ribosomal_L27; 1.
DR PRINTS; PR00063; RIBOSOMAL_L27.
DR ProDom; PD003114; Ribosomal_L27; 1.

```

```

DR TIGRFAMs; TIGR00062; L27; 1.
DR PROSITE; PS00831; RIBOSOMAL_L27; 1.
KW Complete proteome.
SQ SEQUENCE 95 AA; 10464 MW; 735D951C94B7A730 CRC64;

Query Match 100.0%; Score 21; DB 16; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 91 SVIAK 95

RESULT 4
Q8P4C6
ID Q8P4C6 PRELIMINARY; PRT; 98 AA.
AC Q8P4C6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein XCC3784.
GN XCC3784.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Canarotte G., Canavan F., Cardozo J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012499; AAM43030.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; P:receptor activity; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF01476; LysM; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 98 AA; 10593 MW; 0DB6218EB6AFA60 CRC64;

Query Match 100.0%; Score 21; DB 16; Length 98;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 56 SVIAK 60

RESULT 5
Q7YUB8

```

```

ID Q7YUB8 PRELIMINARY; PRT; 107 AA.
AC Q7YUB8;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glutaredoxin.
GN GLX-1.
OS Aphelenchus avenae (Mycophagous nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidea;
OC Aphelenchidae; Aphelenchus.
OX NCBI_TaxID=70226;
RN [1]
RN SEQUENCE FROM N.A.
RA Browne J.A., Goyal K., Tunnacliffe A., Burnell A.;
RT "Expression of a glutaredoxin gene induced by desiccation and
RT oxidative stress in the anhydrobiotic nematode Aphelenchus avenae.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY340999; AAQ20895.1; -.
SQ SEQUENCE 107 AA; 11614 MW; CB0396A67FEC9C32 CRC64;

Query Match 100.0%; Score 21; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 10 SVIAK 14

RESULT 6
Q922Z7 PRELIMINARY; PRT; 111 AA.
ID Q922Z7;
AC Q922Z7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Similar to DNA polymerase beta.
GN POLB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006681; AAH06681.1; -.
DR MGD; MGI:97740; Polb.
DR GO; GO:0006916; P:anti-apoptosis; IMP.
DR GO; GO:0008220; P:necrosis; IMP.
DR InterPro; IPR003583; HHH_1.
DR SMART; SM00278; HHH1; 1.
SQ SEQUENCE 111 AA; 12247 MW; E81BBACDF1A3B44F CRC64;

Query Match 100.0%; Score 21; DB 11; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 44 SVIAK 48

RESULT 7
Q8HAJ2 PRELIMINARY; PRT; 126 AA.
ID Q8HAJ2;
AC Q8HAJ2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Antitermination protein Q.
GN Q.
OS Bacteriophage LC159.

```

```

OC Viruses.
OX NCBI_TaxID=210928;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=159;
RA Muniesa M., Jofre J.;
RT "Variability of shiga converting bacteriophages in E. coli O157:H7
RT strains of human origin isolated from the same outbreak.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF548456; AAN59919.1; -.
SQ SEQUENCE 126 AA; 14230 MW; B8F1776A0329F55A CRC64;

Query Match 100.0%; Score 21; DB 9; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 3 SVIAK 7

RESULT 8
Q8C8J3 PRELIMINARY; PRT; 126 AA.
ID Q8C8J3;
AC Q8C8J3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical S-adenosyl-L-methionine-dependent methyltransferases
DE structure containing protein.
GN 4732479N06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK046928; BAC32921.1; -.
DR MGD; MGI:2442530; 4732479N06RIK.
KW Hypothetical protein.
SQ SEQUENCE 126 AA; 14568 MW; 0AB92B67189578CD CRC64;

Query Match 100.0%; Score 21; DB 11; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 45 SVIAK 49

RESULT 9
O17211 PRELIMINARY; PRT; 132 AA.
ID O17211;
AC O17211;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE C01B12.7 protein.
GN C01B12.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.

```

RC STRAIN=Bristol N2; PubMed=7906398;
 RX MEDLINE=94150718; Anderson K., Baynes C., Berks M., Coulson A.,
 RA Willson R., Ainscough R., Connell M., Copey T., Cooper J., Fulton L.,
 RA Bonfield J., Burton J., Durbin R., Favello A., Fulton L.,
 RA Craxton M., Dear S., Du Z., Durbin R., Hillier L., Jier M., Johnston L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Smaldon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinscock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF025458; AAB70973.1; -.
 DR PIR; T32373; T32373.
 DR WormPep; C01812.7; CE07795.
 SQ SEQUENCE 132 AA; 15750 MW; A2C8BA7465940DF2 CRC64;

Query Match 100.0%; Score 21; DB 5; Length 132;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5

Db 25 SVIAK 29

RESULT 10

Q8SC48 PRELIMINARY; PRT; 135 AA.
 AC Q8SC48;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Stx2 converting bacteriophage I.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
 OC Lambda-like viruses.
 OX NCBI_TaxID=180816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Stx2 phage-I;
 RA Sato T., Shimizu T., Watarai M., Kobayashi M., Kano S., Hamabata T.,
 RA Yamasaki S., Takeda Y.;
 RT "Genomic sequence of Shiga toxin 2-converting phage isolated from
 RT Escherichia coli O157:H7 Okayama strain and comparison with other
 RT Shiga toxin 2-converting phages.";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF004402; BAB87967.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB_boxC.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 135 AA; 16106 MW; 15A614C2A739178C CRC64;

Query Match 100.0%; Score 21; DB 9; Length 135;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SVIAK 5
 Db 100 SVIAK 104
 RESULT 11
 Q8E7B3 PRELIMINARY; PRT; 141 AA.
 ID Q8E7B3;
 AC Q8E7B3;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN GBS0242.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NM316 / Serotype III;
 RX MEDLINE=22242508; PubMed=12354221;
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
 RA Kunst F.;
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease.";
 RL Mol. Microbiol. 45:1499-1513 (2002).
 DR EMBL; AL766844; CAD45887.1; -.
 DR Sagalib; gbs0242; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 141 AA; 15937 MW; 924D2B86930763F5 CRC64;

Query Match 100.0%; Score 21; DB 16; Length 141;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5

Db 21 SVIAK 25

RESULT 12

Q89P48 PRELIMINARY; PRT; 141 AA.
 ID Q89P48;
 AC Q89P48;
 DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE BL13635 protein.
 GN BL13635.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iiguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005948; BAC48900.1; -.
 KW Complete proteome.
 SQ SEQUENCE 141 AA; 14457 MW; 9F10019F39AD214B CRC64;

Query Match 100.0%; Score 21; DB 16; Length 141;

```
Best Local Similarity 100.0%; Pred. No. 4.3e+02; Indels 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0;

QY 1 SVIAK 5
DB 56 SVIAK 60

RESULT 13
Q8VK36 PRELIMINARY; PRT; 152 AA.
AC Q8VK36;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE 4-hydroxyphenylpyruvate dioxygenase C terminal domain containing
DE protein.
GN MT1364.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uutterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007009; AAK45627.1; -.
DR TIGR; MT1364; -.
DR GO; GO:0016702; F:oxidoreductase activity, acting on single d. . .; IEA.
DR InterPro; IPR004360; Gly_bleo_diox.
DR Pfam; PF00903; Glyoxalase; 1.
KW Dioxygenase; Pyruvate.
SQ SEQUENCE 152 AA; 16626 MW; 60B64662DC2B343D CRC64;

Query Match 100.0%; Score 21; DB 16; Length 152;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
DB 81 SVIAK 85

RESULT 14
Q7U080 PRELIMINARY; PRT; 152 AA.
AC Q7U080
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN MB1357C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., MONSENPE C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
```

```
DR EMBL; BX248338; CAD94218.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 16626 MW; 60B64662DC2B343D CRC64;

Query Match 100.0%; Score 21; DB 16; Length 152;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
DB 81 SVIAK 85

RESULT 15
Q9XJQ6 PRELIMINARY; PRT; 161 AA.
AC Q9XJQ6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Q protein.
GN Q.
OS Bacteriophage 21.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=10743;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20092464; PubMed=10628842;
RA Karch H., Schmidt H., Janetzki-Mittmann C., Scheef J., Kroeger M.;
RT "Shiga toxins even when different are encoded at identical positions
RT in the genomes of related temperate bacteriophages.";
RL Mol. Gen. Genet. 262:600-607(1999).
DR EMBL; AJ237660; CAB39993.1; -.
SQ SEQUENCE 161 AA; 18497 MW; A1124675BE0F5896 CRC64;

Query Match 100.0%; Score 21; DB 9; Length 161;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
DB 37 SVIAK 41

Search completed: August 12, 2004, 06:19:32
Job time : 13.5512 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.89549 Seconds

(without alignments)
745.314 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAX 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	5	3	AAY97147 Pigment p
2	21	100.0	5	5	Abb99061 N-termina
3	21	100.0	10	4	Aag87969 Saccharom
4	21	100.0	10	4	Aag87968 Saccharom
5	21	100.0	13	5	Abp70008 Colour Fa
6	21	100.0	16	5	Abb99073 N-termina
7	21	100.0	16	5	Abb99074 N-termina
8	21	100.0	16	5	Abb99072 N-termina
9	21	100.0	17	3	AAY97148 Pigment p
10	21	100.0	18	3	AAY97151 Pigment p
11	21	100.0	25	3	AAY97152 Pigment p
12	21	100.0	50	4	Aab62002 C-mvb pro
13	21	100.0	73	4	Abg14062 Novel hum
14	21	100.0	73	5	Abp34419 Human iso
15	21	100.0	75	6	Abp77483 N. gonorr
16	21	100.0	76	4	Abb12172 Human MEC
17	21	100.0	76	4	Abg16073 Novel hum
18	21	100.0	78	4	Abg16074 Novel hum
19	21	100.0	81	5	Abp39391 Staphyloc
20	21	100.0	89	3	AAG01464 Human sec
21	21	100.0	91	4	Abg16071 Novel hum
22	21	100.0	102	2	Aay29549 Human lun
23	21	100.0	102	3	Aab44472 Human lun
24	21	100.0	102	4	Aae13814 Human lun
25	21	100.0	102	7	Add66504 Human lun

ALIGNMENTS

RESULT 1

AAY97147

ID AAY97147 standard; peptide; 5 AA.

XX AC

XX AAY97147;

XX AC

XX 04-DEC-2000 (first entry)

XX DT

XX XX

XX DE Pigment protein from coral tissue N-terminal peptide 1.

XX XX

XX KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;

XX KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;

XX UV filter.

XX XX

XX OS Acropora aspera.

XX OS Acropora horrida.

XX OS Montipora caliculata.

XX OS Montipora monasteriata.

XX OS Porites murrayensis.

XX OS Porites lobata.

XX XX

XX PN WO200046233-Al.

XX XX

XX PD 10-AUG-2000.

XX XX

XX PF 02-FEB-2000; 2000WO-AU0000056.

XX XX

XX PR 02-FEB-1999; 99AU-00008463.

XX XX

XX PA (UNSY) UNIV SYDNEY.

XX XX

XX PI Hoegh-Guldberg O, Dove S;

XX XX

XX DR WPI; 2000-532892/48.

XX XX

XX PT Novel pigment protein derived from corals capable of emitting fluorescence upon irradiation by incident light useful as tissue marker, fluorescent marker or general dyestuff.

XX PT

XX XX

XX PS Claim 3; Page 42; 49pp; English.

XX XX

XX CC The N-terminal peptides shown in AAY97147-48 are from pigment protein from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon irradiation by incident light whose maximal absorbance is in the range of 320-600 nm and a maximal fluorescence emission is in the range of 300-700 nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to follow gene expression in transformed tissues) or general dyestuff (all claimed). PPCT may also be used in sunscreen formulations or UV filters

XX CC

```

CC (both claimed)
XX
SQ Sequence 5 AA;
  Query Match      100.0%; Score 21; DB 3; Length 5;
  Best Local Similarity 100.0%; Pred. No. 1.4e+06;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SVIAK 5
  |||||
  1 SVIAK 5

RESULT 2
ABB99061
ID ABB99061 standard; peptide; 5 AA.
AC ABB99061;
DT 22-JAN-2003 (first entry)
XX
DE N-terminal amino acid sequence of a CFM #1.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
XX WO200270703-A2.
XX
XX 12-SEP-2002.
XX
XX 01-MAR-2002; 2002WO-GB000928.
XX
XX 02-MAR-2001; 2001US-0273227P.
XX
XX 21-MAR-2001; 2001AU-00003874.
XX
XX 15-OCT-2001; 2001US-0329816P.
XX
XX (NUFA-) NUFARM LTD.
XX
XX (UQUU) UNIV QUEENSLAND.
XX
XX (JONE/) JONES E L.
XX
XX Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
XX Hough-Guldberg IO, Prescott M;
XX
XX WPI; 2002-740765/80.
XX
XX Novel color-facilitating molecule for producing a biomatrix, has a
XX polypeptide which alone/along with molecules imparts altered visual
XX characteristics to cells in the absence of excitation by extraneous non-
XX white light.
XX
XX Claim 3; Page 278; 510pp; English.
XX
XX The invention relates to an isolated colour-facilitating molecule (CFM)
XX comprising a polypeptide which, in a cell, alone or together with one or
XX more other molecules imparts an altered visual characteristic to the cell
XX when visualised by a human eye in the absence of excitation by extraneous
XX non-white light or particle emission. CFMs are useful for producing a
XX transgenic animal which exhibits a novel colour e.g. sheep with blue or
XX red coloured fleece. They are useful for producing coloured plant
XX extracts, e.g. flavouring, beverage or juice or colouring agent. Other
XX uses include transducing or intensifying an image, providing additional
XX light for growing phototropic organisms e.g. algae and/or corals, for
XX coating materials that experience UV damage e.g. plastics and car
XX upholstery. CFMs are useful in the flower industry, in the development of
XX new varieties of flowering plants. Other contemplated uses include,
XX expression markers, general reporter molecules, photon traps, UV sinks or
XX in sunscreens. CFMs modify visible colour in edible and/or ornamental
XX fungal species, and in fruits and vegetables to enhance their
XX marketability. CFMs embedded in a gel matrix improve image quality in

```

```

CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC current sequence represents the N-terminal amino acid sequence of a
CC colour-facilitating molecule (CFM)
XX
SQ Sequence 5 AA;
  Query Match      100.0%; Score 21; DB 5; Length 5;
  Best Local Similarity 100.0%; Pred. No. 1.4e+06;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SVIAK 5
  |||||
  1 SVIAK 5

RESULT 3
AAG87969
ID AAG87969 standard; peptide; 10 AA.
XX
AC AAG87969;
XX
DT 11-SEP-2001 (first entry)
XX
DE Saccharomyces cerevisiae peptide, SEQ ID NO: 2918.
XX
KW Saccharomyces cerevisiae; complementary peptide; peptide identification;
KW drug discovery; drug design.
XX
XX Saccharomyces cerevisiae.
XX
XX WO200142276-A1.
XX
XX 14-JUN-2001.
XX
XX 13-DEC-2000; 2000WO-GB004773.
XX
XX 13-DEC-1999; 99GB-00029471.
XX
XX (PROT-) PROTEOM LTD.
XX
XX Roberts GW, Heal JR;
XX
XX WPI; 2001-367863/38.
XX
XX Identifying complementary peptides by analysis of protein and nucleotide
XX sequence databases, useful in drug design.
XX
XX Example 5; Page 432; 488pp; English.
XX
XX The invention relates to the identification of complementary peptides by
XX analysis of protein and nucleotide sequence databases from higher
XX eukaryotic genomes, excluding human and plants. The specific
XX complementary peptides interact with their relevant target proteins
XX encoded in the eukaryote genome. The peptides may be used as reagents and
XX drugs for drug discovery and as lead ligands for drug design and
XX development. The present sequence is a complementary peptide from
XX Saccharomyces cerevisiae
XX
XX Sequence 10 AA;
  Query Match      100.0%; Score 21; DB 4; Length 10;
  Best Local Similarity 100.0%; Pred. No. 25;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 6 SVIAK 10
  |||||
  6 SVIAK 10

RESULT 4
AAG87968
ID AAG87968 standard; peptide; 10 AA.

```


XX AAG87968;
 XX AC
 XX 11-SEP-2001 (first entry)
 XX
 XX Saccharomyces cerevisiae peptide, SEQ ID NO: 2917.
 XX DE
 XX Saccharomyces cerevisiae;
 XX KW drug discovery; drug design.
 XX KW
 XX OS
 XX Saccharomyces cerevisiae.
 XX PN WO200142276-A1.
 XX PD 14-JUN-2001.
 XX PF 13-DEC-2000; 2000WO-GB004773.
 XX PR 13-DEC-1999; 99GB-00029471.
 XX PA (PROT-) PROTEOM LTD.
 XX PI Roberts GW, Heal JR;
 XX WPI; 2001-367863/38.
 XX PT Identifying complementary peptides by analysis of protein and nucleotide
 XX PT sequence databases, useful in drug design.
 XX PS Example 5; Page 432; 489pp; English.
 XX CC The invention relates to the identification of complementary peptides by
 CC analysis of protein and nucleotide sequence databases from higher
 CC eukaryotic genomes, excluding human and plants. The specific
 CC complementary peptides interact with their relevant target proteins
 CC encoded in the eukaryote genome. The peptides may be used as reagents and
 CC drugs for drug discovery and as lead ligands for drug design and
 CC development. The present sequence is a complementary peptide from
 CC Saccharomyces cerevisiae
 XX
 SQ Sequence 10 AA;
 Query Match 100.0%; Score 21; DB 4; Length 10;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVIAK 5
 Db 6 SVIAK 10
 RESULT 5
 ID ABP70008 standard; peptide; 13 AA.
 XX AC
 XX ABP70008;
 XX
 DT 06-AUG-2003 (revised)
 DT 22-JAN-2003 (first entry)
 XX
 XX Colour Facilitating molecule (CFM) related sequence #SEQ ID 184.
 XX DE
 XX Colour facilitating molecule; CFM; green fluorescent protein; GFP;
 KW chromophore; biomatrix; transgenic animal; colouring agent;
 KW flower industry; expression marker; reporter molecule; photon trap;
 KW UV sink; sunscreen.
 XX Pavona decussata.
 XX OS
 XX WO200270703-A2.
 XX PD 12-SEP-2002.
 XX XX

PF 01-MAR-2002; 2002WO-GB000928.
 XX
 XX 02-MAR-2001; 2001US-0273227P.
 PR 21-MAR-2001; 2001AU-00003874.
 PR 15-OCT-2001; 2001US-0329816P.
 XX
 XX (NUFA-) NUFARM LTD.
 PA (UYOU) UNIV QUEENSLAND.
 PA (JONE/) JONES E L.
 XX
 XX Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
 PI Hoegh-Guldberg IO, Prescott M;
 PI WPI; 2002-740765/80.
 XX
 XX Novel color-facilitating molecule for producing a biomatrix, has a
 PT polypeptide which alone/along with molecules imparts altered visual
 PT characteristics to cells in the absence of excitation by extraneous non-
 PT white light.
 XX
 XX Claim 5; Page 473; 510pp; English.
 PS
 XX The invention relates to an isolated colour-facilitating molecule (CFM)
 CC comprising a polypeptide which, in a cell, alone or together with one or
 CC more other molecules imparts an altered visual characteristic to the cell
 CC when visualised by a human eye in the absence of excitation by extraneous
 CC non-white light or particle emission. CFMs are useful for producing a
 CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
 CC red coloured fleece. They are useful for producing coloured plant
 CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
 CC uses include transducing or intensifying an image, providing additional
 CC light for growing phototropic organisms e.g. algae and/or corals, for
 CC coating materials that experience UV damage e.g. plastics and car
 CC upholstery. CFMs are useful in the flower industry, in the development of
 CC new varieties of flowering plants. Other contemplated uses include,
 CC expression markers, general reporter molecules, photon traps, UV sinks or
 CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
 CC fungal species, and in fruits and vegetables to enhance their
 CC marketability. CFMs embedded in a gel matrix improve image quality in
 CC situations of distorted light spectra (Biomatrix). The first all-protein
 CC chromophore to be isolated was Green Fluorescent protein (GFP). The
 CC sequences given in records ABP6924-ABP70048 represent CFM related amino
 CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 13 AA;
 Query Match 100.0%; Score 21; DB 5; Length 13;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVIAK 5
 Db 1 SVIAK 5
 RESULT 6
 ID ABB99073 standard; peptide; 16 AA.
 XX AC
 XX ABB99073;
 XX
 DT 22-JAN-2003 (first entry)
 XX
 XX N-terminal amino acid sequence of a CFM #13.
 DE
 XX Colour facilitating molecule; CFM; green fluorescent protein; GFP;
 KW chromophore; biomatrix; transgenic animal; colouring agent;
 KW flower industry; expression marker; reporter molecule; photon trap;
 KW UV sink; sunscreen.
 XX Unidentified.
 XX OS
 XX WO200270703-A2.
 XX PN

```

XX PD 12-SEP-2002.
XX PF 01-MAR-2002; 2002WO-GB000928.
XX PR 02-MAR-2001; 2001US-02732227P.
XX PR 21-MAR-2001; 2001AU-00003874.
XX PR 15-OCT-2001; 2001US-0329816P.
XX PA (NUFA-) NUFARM LTD.
XX PA (UYOU ) UNIV QUEENSLAND.
XX PA (JONE/) JONES E L.
XX PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
XX PI Hoegh-Guldberg IO, Prescott M;
XX DR WPI; 2002-740765/80.
XX Novel color-facilitating molecule for producing a biomatrix, has a
XX polypeptide which alone/along with molecules imparts altered visual
XX characteristics to cells in the absence of excitation by extraneous non-
XX white light.
XX Claim 4; Page 281; 510pp; English.
XX The invention relates to an isolated colour-facilitating molecule (CFM)
XX comprising a polypeptide which, in a cell, alone or together with one or
XX more other molecules imparts an altered visual characteristic to the cell
XX when visualised by a human eye in the absence of excitation by extraneous
XX non-white light or particle emission. CFMs are useful for producing a
XX transgenic animal which exhibits a novel colour e.g. sheep with blue or
XX red coloured fleece. They are useful for producing coloured plant
XX extracts, e.g. flavouring, beverage or juice or colouring agent. Other
XX uses include transducing or intensifying an image, providing additional
XX light for growing phototropic organisms e.g. algae and/or corals, for
XX coating materials that experience UV damage e.g. plastics and car
XX upholstery. CFMs are useful in the flower industry, in the development of
XX new varieties of flowering plants. Other contemplated uses include,
XX expression markers, general reporter molecules, photon traps, UV sinks or
XX fungal species, and in fruits and vegetables to enhance their
XX marketability. CFMs embedded in a gel matrix improve image quality in
XX situations of distorted light spectra (biomatrix). The first all-protein
XX chromophore to be isolated was Green Fluorescent protein (GFP). The
XX current sequence represents the N-terminal amino acid sequence of a
XX colour-facilitating molecule (CFM)
XX SQ Sequence 16 AA;
XX Query Match 100.0%; Score 21; DB 5; Length 16;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 SVIAK 5
XX Db |||||
XX 1 SVIAK 5
XX RESULT 7
XX ABB99074
XX ID ABB99074 standard; peptide; 16 AA.
XX AC ABB99074;
XX XX
XX DT 22-JAN-2003 (first entry)
XX XX
XX DE N-terminal amino acid sequence of a CFM #14.
XX XX
XX KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
XX KW chromophore; biomatrix; transgenic animal; colouring agent;
XX KW flower industry; expression marker; reporter molecule; photon trap;
XX KW UV sink; sunsreen.
XX XX

```

```

OS Unidentified.
XX Key Location/Qualifiers
XX FH Misc-difference 10
XX FT /label= Xaa
XX FT /note= "Xaa is any amino acid except lys"
XX FT Misc-difference 11
XX FT /label= Xaa
XX FT /note= "Xaa is any amino acid except Val"
XX FT Misc-difference 13
XX FT /label= Xaa
XX FT /note= "Xaa is any amino acid except Met"
XX XX
XX WO200270703-A2.
XX PN
XX XX
XX PD 12-SEP-2002.
XX PF 01-MAR-2002; 2002WO-GB000928.
XX PR 02-MAR-2001; 2001US-02732227P.
XX PR 21-MAR-2001; 2001AU-00003874.
XX PR 15-OCT-2001; 2001US-0329816P.
XX XX
XX (NUFA-) NUFARM LTD.
XX PA (UYOU ) UNIV QUEENSLAND.
XX PA (JONE/) JONES E L.
XX PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
XX PI Hoegh-Guldberg IO, Prescott M;
XX DR WPI; 2002-740765/80.
XX Novel color-facilitating molecule for producing a biomatrix, has a
XX polypeptide which alone/along with molecules imparts altered visual
XX characteristics to cells in the absence of excitation by extraneous non-
XX white light.
XX Claim 4; Page 282; 510pp; English.
XX The invention relates to an isolated colour-facilitating molecule (CFM)
XX comprising a polypeptide which, in a cell, alone or together with one or
XX more other molecules imparts an altered visual characteristic to the cell
XX when visualised by a human eye in the absence of excitation by extraneous
XX non-white light or particle emission. CFMs are useful for producing a
XX transgenic animal which exhibits a novel colour e.g. sheep with blue or
XX red coloured fleece. They are useful for producing coloured plant
XX extracts, e.g. flavouring, beverage or juice or colouring agent. Other
XX uses include transducing or intensifying an image, providing additional
XX light for growing phototropic organisms e.g. algae and/or corals, for
XX coating materials that experience UV damage e.g. plastics and car
XX upholstery. CFMs are useful in the flower industry, in the development of
XX new varieties of flowering plants. Other contemplated uses include,
XX expression markers, general reporter molecules, photon traps, UV sinks or
XX fungal species, and in fruits and vegetables to enhance their
XX marketability. CFMs embedded in a gel matrix improve image quality in
XX situations of distorted light spectra (biomatrix). The first all-protein
XX chromophore to be isolated was Green Fluorescent protein (GFP). The
XX current sequence represents the N-terminal amino acid sequence of a
XX colour-facilitating molecule (CFM)
XX SQ Sequence 16 AA;
XX Query Match 100.0%; Score 21; DB 5; Length 16;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 SVIAK 5
XX Db |||||
XX 1 SVIAK 5
XX RESULT 8

```

ABB99072
ID ABB99072 standard; peptide; 16 AA.

XX AC ABB99072;
XX DT 22-JAN-2003 (first entry)
XX DE N-terminal amino acid sequence of a CFM #12.
XX KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
XX KW chromophore; biomatrix; transgenic animal; colouring agent;
XX KW flower industry; expression marker; reporter molecule; photon trap;
XX KW UV sink; sunsreen.
XX OS Unidentified.
XX PN WO200270703-A2.
XX PD 12-SEP-2002.
XX PF 01-MAR-2002; 2002WO-GB000928.
XX PR 02-MAR-2001; 2001US-0273227P.
XX PR 21-MAR-2001; 2001AU-00003874.
XX PR 15-OCT-2001; 2001US-0329816P.
XX PA (NUFA-) NUFARM LTD.
XX PA (UYQU) UNIV QUEENSLAND.
XX PA (JONE/) JONES E L.
XX PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
XX PI Hoegh-Guldberg IO, Prescott M;
XX DR WPI; 2002-740765/80.
XX PT Novel color-facilitating molecule for producing a biomatrix, has a
XX PT polypeptide which alone/along with molecules imparts altered visual
XX PT characteristics to cells in the absence of excitation by extraneous non-
XX PT white light.

XX PS Claim 4; Page 281; 510pp; English.

XX CC The invention relates to an isolated colour-facilitating molecule (CFM)
XX CC comprising a polypeptide which, in a cell, alone or together with one or
XX CC more other molecules imparts an altered visual characteristic to the cell
XX CC when visualised by a human eye in the absence of excitation by extraneous
XX CC non-white light or particle emission. CFMs are useful for producing a
XX CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
XX CC red coloured fleece. They are useful for producing coloured plant
XX CC extracts, e.g. flavoured, beverage or juice or colouring agent. Other
XX CC uses include transducing or intensifying an image, providing additional
XX CC light for growing phototropic organisms e.g. algae and/or corals, for
XX CC coating materials that experience UV damage e.g. plastics and car
XX CC upholstery. CFMs are useful in the flower industry, in the development of
XX CC new varieties of flowering plants. Other contemplated uses include,
XX CC expression markers, general reporter molecules, photon traps. UV sinks or
XX CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
XX CC fungal species, and in fruits and vegetables to enhance their
XX CC marketability. CFMs embedded in a gel matrix improve image quality in
XX CC situations of distorted light spectra (biomatrix). The first all-protein
XX CC chromophore to be isolated was Green Fluorescent protein (GFP). The
XX CC current sequence represents the N-terminal amino acid sequence of a
XX CC colour-facilitating molecule (CFM)

XX SQ Sequence 16 AA;

Query Match 100.0%; Score 21; DB 5; Length 16;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 1 SVIAK 5

RESULT 9

AA97148
ID AAY97148 standard; peptide; 17 AA.

XX AC AAY97148;
XX DT 04-DEC-2000 (first entry)
XX DE Pigment protein from coral tissue N-terminal peptide 2.
XX KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
XX KW tissue marker; fluorescent marker; dyestuff; sunsreen; ultra violet;
XX KW UV filter.

XX OS Acropora horrida.

XX PN WO200046233-A1.

XX PD 10-AUG-2000.

XX PF 02-FEB-2000; 2000WO-AU0000056.

XX PR 02-FEB-1999; 99AU-00008463.

XX PA (UNSY) UNIV SYDNEY.

XX PI Hoegh-Guldberg O, Dove S;

XX DR WPI; 2000-532892/48.

XX PT Novel pigment protein derived from corals capable of emitting
XX PT fluorescence upon irradiation by incident light useful as tissue marker,
XX PT fluorescent marker or general dyestuff.

XX PS Claim 4; Page 42; 49pp; English.

XX CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
XX CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
XX CC irradiation by incident light whose maximal absorbance is in the range of
XX CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
XX CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
XX CC follow gene expression in transformed tissues) or general dyestuff (all
XX CC claimed). PPCT may also be used in sunsreen formulations or UV filters
XX CC (both claimed)

XX SQ Sequence 17 AA;

Query Match 100.0%; Score 21; DB 3; Length 17;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 1 SVIAK 5

RESULT 10

AA97151
ID AAY97151 standard; peptide; 18 AA.

XX AC AAY97151;

XX DT 04-DEC-2000 (first entry)

XX DE Pigment protein from coral tissue N-terminal peptide 3.

XX KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
XX KW tissue marker; fluorescent marker; dyestuff; sunsreen; ultra violet;
XX KW UV filter.

XX OS Acropora aspera.

CC homology studies with hcdc5 protein
 XX
 SQ Sequence 50 AA;
 Query Match 100.0%; Score 21; DB 4; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVIAK 5
 Db 26 SVIAK 30
 |||||
 RESULT 13
 ABG14062
 ID ABG14062 standard; protein; 73 AA.
 XX
 AC ABG14062;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #14053.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 FN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS78249.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 44421; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 73 AA;
 Query Match 100.0%; Score 21; DB 4; Length 73;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVIAK 5
 Db 24 SVIAK 28
 |||||
 RESULT 14
 ABP34419
 ID ABP34419 standard; protein; 73 AA.
 XX
 AC ABP34419;
 XX
 DT 09-JUL-2002 (first entry)
 XX
 DE Human isomerase-like ORF3392 protein, SEQ ID NO:6784.
 XX
 KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 XX
 OS Homo sapiens.
 XX
 XX WO200190366-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001WO-US017076.
 XX
 PR 24-MAY-2000; 2000US-0206690P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Leach MD, Shinkets RA;
 XX
 DR WPI; 2002-106200/14.
 DR N-PSDB; ABN78445.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation.
 XX
 PS Claim 10; Page 1943; 2508pp; English.
 XX
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 12, 2004, 06:19:43 ; Search time 10.5225 Seconds
(without alignments)
149.169 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	50	US-09-757-049A-6	Sequence 6, Appli
2	21	100.0	52	US-09-912-962-14	Sequence 14, Appl
3	21	100.0	55	US-10-424-599-235707	Sequence 235707,
4	21	100.0	63	US-10-437-963-201620	Sequence 201620,
5	21	100.0	66	US-10-424-599-281604	Sequence 281604,
6	21	100.0	72	US-10-424-599-259733	Sequence 259733,
7	21	100.0	73	US-09-864-408A-6784	Sequence 6784, Ap
8	21	100.0	76	US-10-276-774-2542	Sequence 2542, Ap
9	21	100.0	102	US-09-738-973-196	Sequence 196, App
10	21	100.0	102	US-09-854-133-196	Sequence 196, App
11	21	100.0	107	US-10-144-649A-196	Sequence 196, App
12	21	100.0	107	US-10-424-599-200366	Sequence 200366,
13	21	100.0	138	US-10-424-599-252896	Sequence 252896,
14	21	100.0	142	US-10-425-114-39108	Sequence 39108, A
15	21	100.0	156	US-09-892-398-4	Sequence 4, Appli

16	21	100.0	156	10	US-09-892-398-46	Sequence 46, Appl
17	21	100.0	161	12	US-10-424-599-283302	Sequence 283302,
18	21	100.0	165	12	US-10-424-599-280513	Sequence 280513, A
19	21	100.0	169	16	US-10-369-493-10328	Sequence 10328, A
20	21	100.0	189	12	US-10-437-963-137115	Sequence 137115,
21	21	100.0	189	12	US-10-424-599-236262	Sequence 236262,
22	21	100.0	192	12	US-10-425-114-64027	Sequence 64027, A
23	21	100.0	193	12	US-10-424-599-180029	Sequence 180029, A
24	21	100.0	196	12	US-10-425-114-72020	Sequence 72020, A
25	21	100.0	196	12	US-10-425-114-72021	Sequence 72021, A
26	21	100.0	208	12	US-10-424-599-209489	Sequence 209489,
27	21	100.0	233	14	US-10-156-761-9990	Sequence 9990, Ap
28	21	100.0	242	12	US-10-424-599-252176	Sequence 252176,
29	21	100.0	244	15	US-10-369-493-21085	Sequence 21085, A
30	21	100.0	249	14	US-10-156-275-88	Sequence 88, Appl
31	21	100.0	260	10	US-09-935-338-105	Sequence 105, App
32	21	100.0	260	12	US-10-380-430-9	Sequence 9, Appli
33	21	100.0	263	12	US-10-282-122A-64121	Sequence 64121, A
34	21	100.0	268	12	US-10-282-122A-63475	Sequence 63475, A
35	21	100.0	273	15	US-10-287-226-374	Sequence 374, App
36	21	100.0	283	12	US-10-425-114-53397	Sequence 53397, A
37	21	100.0	283	12	US-10-425-114-71886	Sequence 71886, A
38	21	100.0	283	12	US-10-425-114-72284	Sequence 72284, A
39	21	100.0	319	9	US-09-864-761-34819	Sequence 34819, A
40	21	100.0	320	9	US-09-886-085-121	Sequence 121, App
41	21	100.0	320	10	US-09-804-291-121	Sequence 121, App
42	21	100.0	320	11	US-09-844-861A-6	Sequence 6, Appli
43	21	100.0	320	12	US-10-343-850A-206	Sequence 206, App
44	21	100.0	320	14	US-10-017-161-942	Sequence 942, App
45	21	100.0	320	15	US-10-292-798-816	Sequence 816, App

ALIGNMENTS

RESULT 1

US-09-757-049A-6
; Sequence 6, Application US/09757049A
; Patent No. US20020127702A1
; GENERAL INFORMATION:
; APPLICANT: BERNSTEIN, Harold S.
; APPLICANT: COUGHLIN, Shaun R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING CELL CYCLE
; TITLE OF INVENTION: PROGRESSION
; FILE REFERENCE: UCSF-020/02US
; CURRENT APPLICATION NUMBER: US/09/757,049A
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/156,316
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: US 60/060,688
; PRIOR FILING DATE: 1997-09-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-049A-6

Query Match 100.0%; Score 21; DB 9; Length 50;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5

DB 26 SVIAK 30

RESULT 2

US-09-912-962-14
; Sequence 14, Application US/09912962
; Patent No. US2002007619A1
; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
; Broccoli, Dominique
; Smogorzewska, Agata
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 52
; ADDRESS: KLAUBER & JACKSON
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/912,962
; FILING DATE: 25-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/018,635
; FILING DATE: 04-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: David A. Jackson
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-142 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-912-962-14

Query Match 100.0%; Score 21; DB 9; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 27 SVIAK 31

RESULT 3
US-10-424-599-235707
; Sequence 235707, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 235707
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_54872C.1.pap

US-10-424-599-235707

Query Match 100.0%; Score 21; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 25 SVIAK 29

RESULT 4
US-10-437-963-201620
; Sequence 201620, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201620
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96978C.1.pap
US-10-437-963-201620

Query Match 100.0%; Score 21; DB 16; Length 63;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
Db 22 SVIAK 26

RESULT 5
US-10-424-599-281604
; Sequence 281604, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 281604
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_96309C.1.pap
US-10-424-599-281604

Query Match 100.0%; Score 21; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 57 SVIAK 61

RESULT 6

US-10-424-599-259733
; Sequence 259733, Application US/10424599
; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 259733

; LENGTH: 72

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_76564C.1.pep

US-10-424-599-259733

Query Match 100.0%; Score 21; DB 12; Length 72;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 32 SVIAK 36

RESULT 7

US-09-864-408A-6784

; Sequence 6784, Application US/09864408A

; Publication No. US20040009474A1

; GENERAL INFORMATION:

; APPLICANT: Leach, Martin D.

; APPLICANT: Shimkets, Richard A.

; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod

; FILE REFERENCE: 21402-012

; CURRENT APPLICATION NUMBER: US/09/864,408A

; CURRENT FILING DATE: 2001-05-24

; PRIOR APPLICATION NUMBER: 60/206,690

; PRIOR FILING DATE: 2000-05-24

; NUMBER OF SEQ ID NOS: 9068

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6784

; LENGTH: 73

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (17)..(17)

; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid

US-09-864-408A-6784

Query Match

Best Local Similarity 100.0%; Score 21; DB 11; Length 73;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 64 SVIAK 68

RESULT 8

US-10-276-774-2542

; Sequence 2542, Application US/10276774

; Publication No. US20040053245A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; APPLICANT: Tang, Y, Tom et al

; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 21272-030

; CURRENT APPLICATION NUMBER: US/10/276,774

; CURRENT FILING DATE: 2002-11-18

; PRIOR APPLICATION NUMBER: 09/560,875

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 09/496,914

; PRIOR FILING DATE: 2000-02-03

; NUMBER OF SEQ ID NOS: 2700

; SOFTWARE: Custom

; SEQ ID NO 2542

; LENGTH: 76

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)..(76)

; OTHER INFORMATION: Xaa = any amino acid or nothing

US-10-276-774-2542

Query Match 100.0%; Score 21; DB 12; Length 76;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 24 SVIAK 28

RESULT 9

US-09-738-973-196

; Sequence 196, Application US/09738973

; Patent No. US20020110563A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Henderson, Robert A.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Fling, Steven P.

; APPLICANT: Mohamath, Raodoh

; APPLICANT: Algate, Paul A.

; APPLICANT: Secrist, Heather

; APPLICANT: Indirias, Carol Yoseph

; APPLICANT: Benson, Darin R.

; APPLICANT: Elliot, Mark

; APPLICANT: Mannion, Jane

; APPLICANT: Kalos, Michael D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.475C9

; CURRENT APPLICATION NUMBER: US/09/738,973

; CURRENT FILING DATE: 2000-12-14

; NUMBER OF SEQ ID NOS: 587

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 196

; LENGTH: 102

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-738-973-196

Query Match

Best Local Similarity 100.0%; Score 21; DB 9; Length 102;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 44 SVIAK 48

RESULT 10
US-09-854-133-196
; Sequence 196, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 196
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-196

Query Match 100.0%; Score 21; DB 9; Length 102;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 44 SVIAK 48

RESULT 11
US-10-144-649A-196
; Sequence 196, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 196
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-144-649A-196

Query Match 100.0%; Score 21; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 44 SVIAK 48

RESULT 12
US-10-424-599-200366
; Sequence 200366, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 200366
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(107)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_22957C.1.pep
US-10-424-599-200366

Query Match 100.0%; Score 21; DB 12; Length 107;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 5 SVIAK 9

RESULT 13
US-10-424-599-252896
; Sequence 252896, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 252896
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_70391C.1.pep
US-10-424-599-252896

Query Match 100.0%; Score 21; DB 12; Length 138;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 125 SVIAK 129

RESULT 14
US-10-425-114-39108
; Sequence 39108, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 39108
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700208712_FLI.pap
US-10-425-114-39108
```

```
Query Match      100.0%; Score 21; DB 12; Length 142;
Best Local Similarity 100.0%; Pred.No. 6.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 SVIAK 5
Db      41 SVIAK 45
```

```
RESULT 15
US-09-892-398-4
; Sequence 4, Application US/09892398
; Publication No. US20030028002A1
; GENERAL INFORMATION:
; APPLICANT: Hirai, Hiroshi
;           Sherk, Charles
;           Inoue, Kazushi
;           Bodner, Sarah M.
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; THEREOF
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
;           Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/892,398
; FILING DATE: 27-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/280,590
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-892-398-4
```

```
Query Match      100.0%; Score 21; DB 10; Length 156;
Best Local Similarity 100.0%; Pred.No. 6.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 SVIAK 5
Db      79 SVIAK 83
```

```
Search completed: August 12, 2004, 06:51:20
Job time : 10.5225 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 0.522541 Seconds
(without alignments)
493.990 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	50	US-09-156-316-6	Sequence 6, Appl
2	21	100.0	52	US-08-519-103-14	Sequence 14, Appl
3	21	100.0	52	US-09-018-635-14	Sequence 14, Appl
4	21	100.0	52	US-09-912-962-14	Sequence 14, Appl
5	21	100.0	81	US-09-134-001C-4236	Sequence 4236, Ap
6	21	100.0	102	US-09-370-838-196	Sequence 196, App
7	21	100.0	111	US-09-107-532A-5388	Sequence 5388, Ap
8	21	100.0	156	US-08-928-941D-4	Sequence 4, Appli
9	21	100.0	156	US-08-928-941D-36	Sequence 36, Appl
10	21	100.0	156	US-09-280-590A-4	Sequence 4, Appli
11	21	100.0	156	US-09-280-590A-46	Sequence 46, Appl
12	21	100.0	156	US-09-892-398-4	Sequence 4, Appli
13	21	100.0	156	US-09-892-398-46	Sequence 46, Appl
14	21	100.0	212	US-09-489-039A-12172	Sequence 12172, A
15	21	100.0	244	US-09-107-532A-5886	Sequence 5886, Ap
16	21	100.0	249	US-08-680-726A-88	Sequence 88, Appl
17	21	100.0	249	US-09-092-409-88	Sequence 88, Appl
18	21	100.0	255	US-09-328-352-6414	Sequence 6414, Ap
19	21	100.0	332	US-09-252-991A-24865	Sequence 24865, A
20	21	100.0	342	US-09-381-810A-1	Sequence 1, Appli
21	21	100.0	354	US-09-328-352-7825	Sequence 7825, Ap
22	21	100.0	401	US-09-465-558-70	Sequence 70, Appl
23	21	100.0	436	US-09-543-681A-6760	Sequence 6760, Ap
24	21	100.0	529	US-09-323-998E-23	Sequence 23, Appl
25	21	100.0	529	US-09-323-998E-47	Sequence 47, Appl
26	21	100.0	529	US-09-323-998E-50	Sequence 50, Appl
27	21	100.0	529	US-09-323-998E-51	Sequence 51, Appl

28 21 100.0 598 4 US-09-134-000C-4957 Sequence 4957, Ap
29 21 100.0 713 3 US-09-335-409-11 Sequence 11, Appl
30 21 100.0 713 4 US-09-568-102-11 Sequence 11, Appl
31 21 100.0 713 4 US-09-567-969-11 Sequence 11, Appl
32 21 100.0 713 4 US-09-568-480-11 Sequence 11, Appl
33 21 100.0 713 4 US-09-568-486-11 Sequence 11, Appl
34 21 100.0 713 4 US-09-568-472-11 Sequence 11, Appl
35 21 100.0 713 4 US-09-567-899-11 Sequence 11, Appl
36 21 100.0 1504 4 US-09-328-352-7046 Sequence 7046, Ap
37 20 95.2 32 3 US-08-433-522A-16 Sequence 16, Appl
38 20 95.2 32 3 US-09-135-166-16 Sequence 16, Appl
39 20 95.2 32 3 US-08-942-046-16 Sequence 16, Appl
40 20 95.2 48 3 US-09-107-858-24 Sequence 24, Appl
41 20 95.2 48 4 US-09-579-174-24 Sequence 24, Appl
42 20 95.2 59 1 US-08-485-455D-71 Sequence 71, Appl
43 20 95.2 59 2 US-08-482-130C-71 Sequence 71, Appl
44 20 95.2 59 2 US-08-484-211C-71 Sequence 71, Appl
45 20 95.2 59 3 US-08-906-769-71 Sequence 71, Appl

ALIGNMENTS

RESULT 1

US-09-156-316-6
; Sequence 6, Application US/09156316
; Patent No. 6183961
; GENERAL INFORMATION:
; APPLICANT: Bernstein, Harold S.
; APPLICANT: Coughlin, Shaun R.
; TITLE OF INVENTION: Methods and Compositions for Regulating Cell Cycle
; FILE REFERENCE: UCSF-020/01US
; CURRENT APPLICATION NUMBER: US/09/156,316
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 60/060,688
; EARLIER FILING DATE: 1997-09-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-156-316-6

Query Match 100.0%; Score 21; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SVIAK 5

Db 26 SVIAK 30

RESULT 2

US-08-519-103-14
; Sequence 14, Application US/08519103
; Patent No. 5733730
; GENERAL INFORMATION:
; APPLICANT: deLange, Titia
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KLAUBER & JACKSON
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/519,103
; FILING DATE: 25-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feuzy, Sharon E.
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 600-1-142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-519-103-14

Query Match 100.0%; Score 21; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 27 SVIAK 31

RESULT 3
US-09-018-635-14
; Sequence 14, Application US/09018635
; Patent No. 6297356
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Broccoli, Dominique
; APPLICANT: Smogorzewska, Agata
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KLAUBER & JACKSON
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; APPLICATION NUMBER: US/09/018,635
; ATTORNEY/AGENT INFORMATION:
; NAME: David A. Jackson
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-142 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: peptide
; US-09-018-635-14

Query Match 100.0%; Score 21; DB 3; Length 52;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 27 SVIAK 31

RESULT 4
US-09-912-962-14
; Sequence 14, Application US/09912962
; Patent No. 6586577
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Broccoli, Dominique
; APPLICANT: Smogorzewska, Agata
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KLAUBER & JACKSON
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/912,962
; FILING DATE: 25-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/018,635
; FILING DATE: 04-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: David A. Jackson
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-142 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
; US-09-912-962-14

Query Match 100.0%; Score 21; DB 4; Length 52;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 27 SVIAK 31

RESULT 5
US-09-134-001C-4236
; Sequence 4236, Application US/09134001C
; Patent No. 6380370

```

```
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4236
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4236

Query Match          100.0%; Score 21; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SVIAK 5
Db      77 SVIAK 81

RESULT 6
; US-09-370-838-196
; Sequence 196, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 196
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-370-838-196

Query Match          100.0%; Score 21; DB 4; Length 102;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SVIAK 5
Db      44 SVIAK 48

RESULT 7
; US-09-107-532A-5388
; Sequence 5388, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham

; GENERAL INFORMATION:
; APPLICANT: Hiral, Hiroshi
; APPLICANT: Sherr, Charles
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,941D
; FILING DATE:
```

```
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5388:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...111
; SEQUENCE DESCRIPTION: SEQ ID NO: 5388:
US-09-107-532A-5388

Query Match          100.0%; Score 21; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SVIAK 5
Db      58 SVIAK 62

RESULT 8
; US-08-928-941D-4
; Sequence 4, Application US/08928941D
; Patent No. 6180763
; GENERAL INFORMATION:
; APPLICANT: Hiral, Hiroshi
; APPLICANT: Sherr, Charles
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,941D
; FILING DATE:
```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
US-08-928-941D-4

Query Match 100.0%; Score 21; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 79 SVIAK 83

RESULT 9
US-08-928-941D-36
; Sequence 36, Application US/08928941D
; Patent No. 6180763
; GENERAL INFORMATION:
; APPLICANT: Hiral, Hiroshi
; APPLICANT: Sherr, Charles
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,941D
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; FRAGMENT TYPE:

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
US-09-280-590A-4

Query Match 100.0%; Score 21; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 79 SVIAK 83

RESULT 10
US-09-280-590A-4
; Sequence 4, Application US/09280590A
; Patent No. 6303772
; GENERAL INFORMATION:
; APPLICANT: Hiral, Hiroshi
; APPLICANT: Sherr, Charles
; APPLICANT: Inoue, Kazushi
; APPLICANT: Bodner, Sarah M.
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/280,590A
; FILING DATE: 29-Mar-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
US-09-280-590A-4

Query Match 100.0%; Score 21; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 79 SVIAK 83

RESULT 11

```


US-09-280-590A-46
; Sequence 46, Application US/09280590A
; Patent No. 6303772
; GENERAL INFORMATION:
; APPLICANT: Hirai, Hiroshi
; Sherr, Charles
; Inoue, Kazushi
; Bodner, Sarah M.
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; THEREOF
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/280,590A
; FILING DATE: 29-Mar-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; ORGANISM: Gallus gallus
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-280-590A-46
Query Match 100.0%; Score 21; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVIAK 5
Db |||||
79 SVIAK 83
RESULT 12
US-09-892-398-4
; Sequence 4, Application US/09892398
; Patent No. 6673902
; GENERAL INFORMATION:
; APPLICANT: Hirai, Hiroshi
; Sherr, Charles
; Inoue, Kazushi
; Bodner, Sarah M.
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; THEREOF
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; Floor

Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/892,398
; FILING DATE: 27-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/280,590
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-892-398-4
Query Match 100.0%; Score 21; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVIAK 5
Db |||||
79 SVIAK 83
RESULT 13
US-09-892-398-46
; Sequence 46, Application US/09892398
; Patent No. 6673902
; GENERAL INFORMATION:
; APPLICANT: Hirai, Hiroshi
; Sherr, Charles
; Inoue, Kazushi
; Bodner, Sarah M.
; TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
; THEREOF
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/892,398
; FILING DATE: 27-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/280,590
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; ORGANISM: Gallus gallus
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-892-398-46
Query Match 100.0%; Score 21; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 79 SVIAK 83

RESULT 14
US-09-489-039A-12172
; Sequence 12172, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12172
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12172
Query Match 100.0%; Score 21; DB 4; Length 212;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 36 SVIAK 40

RESULT 15
US-09-107-532A-5886
; Sequence 5886, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5886:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...244
; SEQUENCE DESCRIPTION: SEQ ID NO: 5886:
US-09-107-532A-5886
Query Match 100.0%; Score 21; DB 4; Length 244;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 186 SVIAK 190

Search completed: August 12, 2004, 06:21:04
Job time : 1.52254 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.56762 Seconds
(without alignments)
1043.144 Million cell updates/sec

Title: US-09-890-463-2

Perfect score: 83

Sequence: 1 SVIAKQMTYKVMGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	54.2	398	2 G95956	protein of short-c
2	43	51.8	424	2 T43498	hypothetical prote
3	43	51.8	3770	2 A40889	delta-(l-alpha-am
4	41	49.4	3587	2 I40486	surfactin syntheta
5	41	49.4	3588	2 I40485	surfactin syntheta
6	40	48.2	218	2 E86687	hypothetical prote
7	40	48.2	263	2 C97225	hypothetical prote
8	40	48.2	285	2 F83890	sugar transport sy
9	40	48.2	294	2 G83962	hypothetical prote
10	40	48.2	870	2 T10634	hypothetical prote
11	40	48.2	1086	2 AH2136	microcystin synthe
12	39	47.0	161	2 A40745	odorant receptor (
13	39	47.0	170	2 S64488	regulatory protein
14	39	47.0	277	2 A70233	hypothetical prote
15	39	47.0	337	2 G70394	Plex protein - Aqu
16	39	47.0	346	2 H71473	probable leucine d
17	39	47.0	368	2 F81816	phosphoserine tran
18	39	47.0	368	2 H81059	phosphoserine amin
19	39	47.0	459	2 T29945	hypothetical prote
20	39	47.0	786	2 AG0182	probable membrane
21	39	47.0	846	2 YS3262	hypothetical prote
22	39	47.0	3712	1 YGCEVC	alpha-aminoadipyl-
23	38	45.8	140	2 A69445	hypothetical prote
24	38	45.8	192	2 A10383	probable lipoprote
25	38	45.8	192	2 AH0557	probable lipoprote
26	38	45.8	226	2 D85540	probable polymeras
27	38	45.8	226	2 B64773	lipoprotein yajG p
28	38	45.8	226	2 H90689	probable polymeras
29	38	45.8	227	2 E25973	pertussis toxin ch

30	38	45.8	234	2 F95952	probable membrane-
31	38	45.8	304	2 B36716	protochlorophyllid
32	38	45.8	353	1 KIBEMV	thymidine kinase (
33	38	45.8	554	2 H70011	exo-alpha-1,4-gluc
34	38	45.8	560	2 S38035	probable serine/th
35	38	45.8	635	2 S73358	topoisomerase IV c
36	37	44.6	227	1 F25973	pertussis toxin ch
37	37	44.6	351	2 E86605	leucine dehydrogen
38	37	44.6	351	2 F72020	leucine dehydrogen
39	37	44.6	384	2 G81436	probable serine/th
40	37	44.6	440	2 F84955	UDP-N-acetylmuramo
41	37	44.6	478	1 S03826	UMP synthase - sli
42	37	44.6	552	2 B86899	hypothetical prote
43	37	44.6	563	2 S10176	protochlorophyllid
44	37	44.6	1184	2 T41515	coiled coil protei
45	37	44.6	2560	1 I40457	peptide synthetase

ALIGNMENTS

RESULT 1

G96956

protein of short-chain alcohol dehydrogenase family [imported] - Clostridium acetobutylic

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: G96956

R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: G96956

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-398 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78442.1; PID:gl5023320; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0462

C:Superfamily: Xylella fastidiosa hypothetical protein XF1835

Query Match 54.2%; Score 45; DB 2; Length 398;

Best Local Similarity 47.1%; Pred. No. 3.2;

Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKVMGTV 17

DB 224 SVIGSPRTYKIVREGTI 240

RESULT 2

T43498

hypothetical protein DKFZp586C1021.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000

R:Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z22515

A:Accession: T43498

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-424 <AAA>

A:Cross-references: EMBL:AL33640

A:Experimental source: adult uterus; clone DKFZp586C1021

C:Genetics:

A>Note: DKFZp586C1021.1

Query Match

Best Local Similarity 51.8%; Score 43; DB 2; Length 424;

Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

QY      1 SVIAKQMTYKVMYSGT 16
      |:.|:.|:.|:.|:.|:.|
Db      1 SIVALKNSYEVYFTGT 16

RESULT 3
A40889
delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase - Emericella nidulans
N:Alternate names: ACV synthetase
C:Species: Emericella nidulans, Aspergillus nidulans
C>Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 03-Nov-2000
C:Accession: A40889; S16466
R:MacCabe, A.P.; van Liempt, H.; Palissa, H.; Unkles, S.E.; Riach, M.B.R.; Pfeifer, E.;
J. Biol. Chem. 266, 12646-12654, 1991
A:Title: delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from Aspergillus ni-
thway.
A:Reference number: A40889; MUID:91286299; PMID:2061333
A:Accession: A40889
A:Molecule type: DNA
A:Residues: 1-3770 <MAC>
A:Cross-references: GB:X54853; NID:92318; PIDN:CAA38631.1; PID:g2319
A:Note: the sequence of residues 3129-3148 and the corresponding nucleotide sequence are
C:Genetics:
A:Gene: acvA
C:Superfamily: alpha-aminoadipyl-cysteinyl-valine synthetase; acetate-CoA ligase homolog
C:Keywords: antibiotic biosynthesis; carrier protein; penicillin biosynthesis; phosphop
F:365-831/Domain: acetate-CoA ligase homology <ACL1>
F:848-918/Domain: acyl carrier protein homology <ACP1>
F:1458-1915/Domain: acetate-CoA ligase homology <ACL2>
F:1931-2001/Domain: acyl carrier protein homology <ACP2>
F:2539-3001/Domain: acetate-CoA ligase homology <ACL3>
F:3018-3086/Domain: acyl carrier protein homology <ACP3>
F:882,1965,3050/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match      51.8%; Score 43; DB 2; Length 3770;
Best Local Similarity 50.0%; Pred. No. 71;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 SVIAKQMTYKVMYSGT 16
      |:.|:.|:.|:.|:.|:.|
Db      468 SLTSKQLAYVYTSCT 483

RESULT 4
I40486
surfactin synthetase component II - Bacillus subtilis
N:Alternate names: surfactin synthetase srfA2; surfactin synthetase/competence protein s
N:Contains: acid-amino-acid ligase (EC 6.3.2.-)
C:Species: Bacillus subtilis
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
C:Accession: I40486; S60866; G69718; S46968; S35518; S25658; S34986
R:Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van Sind
Mol. Microbiol. 8, 821-831, 1993
A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A:Reference number: I40486; MUID:93360813; PMID:8355609
A:Accession: I40486
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3587 <RES>
A:Cross-references: EMBL:X70356; NID:g396480; PIDN:CAA49817.1; PID:g396482
A:Experimental source: strain W168 derivative of JH642
R:Hansen, L.W.; Eshuis, H.; Jongbloed, J.; Venema, G.; van Sinderen, D.
Mol. Microbiol. 15, 55-63, 1995
A:Title: A small gene, designated comS, located within the coding region of the fourth a
A:Reference number: S60866; MUID:95272393; PMID:7752896
A:Accession: S60866
A:Molecule type: DNA
A:Residues: 977-1104 <HAM>
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen

```

```

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshiida, K.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69718
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3587 <KUN>
A:Cross-references: GB:Z99105; GB:AL009136; NID:g2632457; PIDN:CAB12143.1; PID:g2632635
A:Experimental source: strain 168
R:Fabret, C.; Quentin, Y.; Giuseppe, A.; Busuttill, J.; Haiech, J.; Denizot, F.
submitted to the EMBL Data Library, March 1993
A:Reference number: S46967
A:Accession: S46968
A:Molecule type: DNA
A:Residues: 1-32, 'F', 34-41, 'G', 43-109, 'D', 111-114, 'G', 116-138, 'V', 140-258, 'W', 260-308, 'A'
1756-1914, 'PK', 1917-2138, 'SR', 2142, 'DSNV', 2146-2444, 'Q', 2446-2712, 'H', 2714-2722, 'H', 272
A:Cross-references: EMBL:X72672; NID:g516358; PIDN:CAA51223.1; PID:g516360
R:Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M.M.; Zuber, P.; Yamane, K.
Nucleic Acids Res. 21, 93-97, 1993
A:Title: Nucleotide sequence of 5' portion of srfA that contains the region required for
A:Reference number: S35517; MUID:93181186; PMID:8441623
A:Accession: S35518
A>Status: significant sequence differences
A:Molecule type: DNA
A:Cross-references: EMBL:D13262; NID:g216345; PID:g216347
A:Experimental source: strain 168 trpC2
R:Borchert, S.; Batil, S.S.; Marahiel, M.A.
FEMS Microbiol. Lett. 92, 175-180, 1992
A:Title: Identification of putative multifunctional peptide synthetase genes using highl
A:Reference number: S25658
A:Accession: S25658
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 168, 'C', 170-171, 281-283, 514-595, 597-647, 'R', 649-679, 'ETL', 683-693, 'DKR', 697,
A:Cross-references: EMBL:X65835; NID:g40202; PIDN:CAA46678.1; PID:g40203
A:Experimental source: strain ATCC 21332
C:Comment: This protein contains several amino acid-activating domains for the synthesis
the amino-terminal region of this protein, appear to be required for the development of
C:Genetics:
A:Gene: srfA2; srfA2
C:Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h
C:Keywords: antibiotic biosynthesis; carrier protein; duplication; ligase; phosphopantet
F:511-951/Domain: acetate-CoA ligase homology <ACL1>
F:968-1035/Domain: acyl carrier protein homology <ACP1>
F:1036-1481/Domain: repeat <RPT1>
F:1542-1995/Domain: acetate-CoA ligase homology <ACL2>
F:2013-2081/Domain: acyl carrier protein homology <ACP2>
F:2082-2529/Domain: repeat <RPT2>
F:2591-3024/Domain: acetate-CoA ligase homology <ACL3>
F:3041-3108/Domain: acyl carrier protein homology <ACP3>
F:999, 2045, 3073/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match      49.4%; Score 41; DB 2; Length 3587;
Best Local Similarity 43.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY      1 SVIAKQMTYKVMYSGT 16
      |:.|:.|:.|:.|:.|:.|
Db      2687 AVTAENLAYMYTSCT 2702

RESULT 5
I40485
surfactin synthetase component I - Bacillus subtilis
N:Alternate names: competence protein srfA; surfactin production protein srfAA; surfact

```

C;Species: Bacillus subtilis
 C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
 C;Accession: I40485; B69718; S35517; A37323; S46367; A43705; S54985
 R;Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Venema, G.; van Sind
 Mol. Microbiol. 8, 821-831, 1993
 A;Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
 A;Reference number: I40485; MUID:93360813; PMID:8355609
 A;Accession: I40485
 A;Status: translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-3588 <RES>
 A;Cross-references: EMBL:X70356; NID:G9396480; PIDN:CAA49816.1; PID:G9396481
 A;Experimental source: strain W168 derivative of JH642
 R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C;Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: B69718
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-3588 <KUN>
 A;Cross-references: GB:Z89105; GB:AL009126; NID:92632457; PIDN:CAB12142.1; PID:G2632634
 A;Experimental source: strain 168
 R;Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M.M.; Zuber, P.; Yamane, K.
 Nucleic Acids Res. 21, 93-97, 1993
 A;Title: Nucleotide sequence of 5' portion of srfA that contains the region required for
 A;Reference number: S35517; MUID:93181186; PMID:8441623
 A;Accession: S35517
 A;Status: nucleic acid sequence not shown; significant sequence differences
 A;Molecule type: DNA
 A;Cross-references: EMBL:D13262; NID:G216345; PID:G216346
 A;Experimental source: strain 168 trpC2
 A;Note: protein sequence not complete, the nucleotide sequence was submitted to the EMBL
 R;Nakano, M.M.; Magnuson, R.; Myers, A.; Curry, J.; Grossman, A.D.; Zuber, P.
 J. Bacteriol. 173, 1770-1778, 1991
 A;Title: srfA is an operon required for surfactin production, competence development, an
 A;Reference number: A37323; MUID:91154134; PMID:1847909
 A;Accession: A37323
 A;Status: nucleic acid sequence not shown; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-46, 'EV', 49-145, 'I', 147-150, 'L', 152, 'AN', 155-280, 'T', 282-307, 'SF', 310-384 <
 A;Cross-references: GB:M59939; NID:G14366
 R;Nakano, M.M.; Xia, L.; Zuber, P.
 J. Bacteriol. 173, 5487-5493, 1991
 A;Title: Transcription initiation region of the srfA operon, which is controlled by the
 A;Reference number: A43705; MUID:91358926; PMID:1715856
 A;Contents: annotation
 R;Fabret, C.; Quentin, Y.; Guiseppe, A.; Busuttil, J.; Haiech, J.; Denizot, F.
 submitted to the EMBL Data Library, March 1993
 A;Reference number: S46967
 A;Accession: S46967
 A;Molecule type: DNA
 A;Residues: 3249-3271, 'A', 3273-3316, 'R', 3318-3451, 'Y', 3453-3483, 'DE', 3486-3487, 'DAGL', 34
 A;Cross-references: EMBL:X72672; NID:G516358; PIDN:CAA51222.1; PID:G516359
 A;Experimental source: strain 168 trpC2
 C;Comment: This enzyme is one of several in the multienzyme complex that synthesizes the
 nce.
 C;Genetics:
 A;Gene: srfAA; srfAI
 A;Note: srfAA is the first gene of the srfA operon
 C;Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein H

C;Keywords: antibiotic biosynthesis; carrier protein; phosphopantetheine; phosphoprotein
 F;507-956/Domain: acetate-CoA ligase homology <ACL1>
 F;574-1042/Domain: acyl carrier protein homology <ACP1>
 F;1043-1488/Domain: repeat <RPT1>
 F;1549-1993/Domain: acetate-CoA ligase homology <ACL2>
 F;2011-2079/Domain: acyl carrier protein homology <ACP2>
 F;2080-2527/Domain: repeat <RPT2>
 F;2589-3025/Domain: acetate-CoA ligase homology <ACL3>
 F;3042-3109/Domain: acyl carrier protein homology <ACP3>
 F;1006,2043,3074/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
 Query Match 49.4%; Score 41; DB 2; Length 3588;
 Best Local Similarity 43.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SVIAKQMTYKVYMSGT 16
 :||: :|:||||
 Db 2688 AVTAENLAYMIYTSQT 2703
 RESULT 6
 E86687
 Hypothetical protein [imported] - Lactococcus lactis subsp. lactis (strain Ifl1403)
 C;Species: Lactococcus lactis subsp. lactis
 C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C;Accession: E86687
 R;Polotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlic
 Genome Res. 11, 731-753, 2001
 A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
 A;Reference number: A86625; MUID:21235186; PMID:11337471
 A;Accession: E86687
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-218 <STO>
 A;Cross-references: GB:AB005176; PID:G12723383; PIDN:AAK04599.1; GSPDB:GN00146
 A;Experimental source: strain IL1403
 C;Genetics:
 A;Gene: yfaA
 Query Match 48.2%; Score 40; DB 2; Length 218;
 Best Local Similarity 70.0%; Pred. NO. 14;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 5 KOMTYKVYMS 14
 :|||||:|
 Db 100 KEMTYKFIIS 109
 RESULT 7
 C97225
 Hypothetical protein CAC2643 [imported] - Clostridium acetobutylicum
 C;Species: Clostridium acetobutylicum
 C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C;Accession: C97225
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeug, Q.; Gibson, R.; Lee,
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 A;Reference number: A96900; MUID:21359325; PMID:21359325
 A;Accession: C97225
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-263 <KUR>
 A;Cross-references: GB:AB001437; PIDN:AAK00590.1; PID:G15025670; GSPDB:GN00168
 A;Experimental source: Clostridium acetobutylicum ATCC824
 C;Genetics:
 A;Gene: CAC2643
 Query Match 48.2%; Score 40; DB 2; Length 263;
 Best Local Similarity 43.8%; Pred. NO. 17;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SVIAKQMTYKVYMSGT 16

A:Accession: A01413
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-161 <RES>
 A:Cross-references: GB:LI4568; NID:q293757; PIDN:AAA39852.1; PID:G
 A:Experimental source: olfactory epithelium
 A>Note: sequence extracted from NCBI backbone (NCBIP:131747)
 C:Superfamily: olfactory receptor OR14
 C:Keywords: G protein-coupled receptor; transmembrane protein
 Query Match 47.0%; Score 39; DB 2; Length 161;
 Best Local Similarity 66.7%; Pred. No. 16;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 IAKQMTYKVYMS 14
| | | | | | | |
Db 5 ICKPLTYKVYMS 16
| | | | | | | |

RESULT 13

S64488
regulatory protein CBP4 precursor - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G7122; protein YGR174c
C:Species: Saccharomyces cerevisiae
C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 21-Jul-2000
C:Accession: S64488; A53928
R:Hebling, U.; Hofmann, B.; Delius, H.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64003
A:Accession: S64488
A:Molecule type: DNA
A:Residues: 1-170 <HEB>
A:Cross-references: EMBL:Z72959; NID:G1323307; PID:e243557; MIPS:YGR174c
R:Crivellone, M.D.
J. Biol. Chem. 269, 21284-21292, 1994
A:Title: Characterization of CBP4, a new gene essential for the expression of ubiquinol-
A:Reference number: A53928; MUID:94342301; PMID:8063753
A:Accession: A53928
A:Molecule type: DNA
A:Residues: 1-64, F', 66-170 <CRI>
A:Cross-references: GB:U10700; NID:G505645; PIDN:AAA61566.1; PID:G505646
C:Genetics:
A:Gene: SGD:CBP4
A:Cross-references: SGD:S0003406; MIPS:YGR174c
A:Map position: 7R
A:Genome: nuclear
C:Keywords: mitochondrion; transmembrane protein
F30-51/Domain: transmembrane #status predicted <TMM>

Query Match 47.0%; Score 39; DB 2; Length 170;
Best Local Similarity 61.5%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKVYMS 13
: | | | | | | | |
Db 10 AVIAKQRYKHYL 22
: | | | | | | | |

RESULT 14

A70233
Hypothetical protein BBG17 - Lyme disease spirochete plasmid G/lp28-2
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 28-Jul-2000
C:Accession: A70233
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: A70233
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-277 <KLE>
A:Cross-references: GB:AEC00786; NID:G2690008; PIDN:AAC66065.1; PID:G2690022; TIGR:BBG17
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid
C:Superfamily: Borrelia burgdorferi hypothetical protein BBG17

Query Match 47.0%; Score 39; DB 2; Length 277;
Best Local Similarity 41.2%; Pred. No. 27;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKVYMSGTV 17
| : | : | : | : | : | : |
Db 70 SIFFKEMAYKMHVFDTV 86
| : | : | : | : | : | : |

RESULT 15

G70394
PlsX protein - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 28-Jul-2003
C:Accession: G70394
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Over-
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: G70394
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-337 <AQF>
A:Cross-references: GB:AEO00723; NID:G2983569; PIDN:AAC07145.1; PID:G2983573; GB:AEO0065;
A:Experimental source: strain VFS
C:Genetics:
A:Gene: plsX
C:Superfamily: phospholipid biosynthesis protein, PlsX type

Query Match 47.0%; Score 39; DB 2; Length 337;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 AKQMTYKVYMSG 15
| : | : | : | : | : |
Db 26 AKELGYKTYLVG 37
| : | : | : | : | : |

Search completed: August 12, 2004, 06:13:50
Job time : 3.56762 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.04508 Seconds
(without alignments)
847.008 Million cell updates/sec

Title: US-09-890-463-2

Perfect score: 83

Sequence: 1 SVIAKQMTYKYVMSGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	43	51.8	3770	1 ACVS EMENI	P27742 emericella
2	41	49.4	3587	1 SRF2_BACSU	Q04747 bacillus su
3	41	49.4	3588	1 SRF1_BACSU	P27206 bacillus su
4	40	48.2	294	1 ENGC_BACHD	Q9k921 bacillus ha
5	40	48.2	697	1 SMZA_SCHGR	O9xzc8 schistocer
6	39	47.0	161	1 OL7A_MOUSE	P34985 mus musculu
7	39	47.0	170	1 CBP4_YEAST	P37267 saccharomyc
8	39	47.0	337	1 PLSX_AQURE	O67186 aquifex aeo
9	39	47.0	368	1 SERC_NEIMA	Q34370 neisseria m
10	39	47.0	368	1 SERC_NEIMB	P57007 neisseria m
11	39	47.0	846	1 SP98_YEAST	P53540 saccharomyc
12	39	47.0	3712	1 ACVS_CEPAC	P25464 cephalospor
13	38	45.8	140	1 VP62_ARCFU	O28710 archaeoglob
14	38	45.8	192	1 YAJG_ECOLI	P36671 escherichia
15	38	45.8	304	1 BCLH_RHOCA	P26237 rhodobacter
16	38	45.8	352	1 KITH_HSVMD	P17653 marek's dis
17	38	45.8	386	1 MTLD_OCEIH	O8en87 oceanobacil
18	38	45.8	560	1 PTK1_YEAST	P36002 saccharomyc
19	38	45.8	635	1 PARE_MYCPN	P78016 mycoplasma
20	37	44.6	440	1 MURD_BUCAT	P57313 buchnera ap
21	37	44.6	444	1 CHLB_ECHLPT	P37824 chlamydomon
22	37	44.6	478	1 PYRS_DICDI	P09556 dictyosteli
23	37	44.6	563	1 CHLB_CHLMO	P17652 chlamydomon
24	37	44.6	2560	1 PPS2_BACSU	P39845 bacillus su
25	36	43.4	134	1 YA7B_HAEIN	Q57425 haemophilus
26	36	43.4	155	1 H2B4_VOLCA	P16868 volvox cart
27	36	43.4	157	1 H2B3_VOLCA	P16867 volvox cart
28	36	43.4	172	1 FAB4_VIBPA	Q87pc5 vibrio para
29	36	43.4	172	1 FAB4_VIBVU	Q8d9h3 vibrio vuln
30	36	43.4	190	1 KCY_THEVO	Q37bv0 thermoplasm
31	36	43.4	232	1 RHOE_HUMAN	Q2730 homo sapien
32	36	43.4	297	1 BCLH_RHOSH	Q9frd6 rhodobacter
33	36	43.4	349	1 IL8A_RAT	P70612 rattus norv

34	36	43.4	379	1 PURK_BACSU	P12045 bacillus su
35	36	43.4	384	1 YH74_VIBCH	Q9kx69 vibrio chol
36	36	43.4	453	1 MPL_HAEIN	P43948 haemophilus
37	36	43.4	489	1 SYE_OCEIH	Q8eu02 oceanobacil
38	36	43.4	517	1 MB13_YEAST	Q9zzw7 saccharomyc
39	36	43.4	1224	1 RPOD_PINTH	P41606 pinus thunb
40	36	43.4	6548	1 EPPL_MOUSE	Q8r0w0 mus musculu
41	35	42.2	60	1 CX3_NAJHA	P01459 naja haje a
42	35	42.2	60	1 CX4_NAJHA	P01461 naja haje a
43	35	42.2	60	1 CX8_NAJHA	P01460 naja haje a
44	35	42.2	100	1 SUI1_THEVO	Q97bw9 thermoplasm
45	35	42.2	171	1 FABA_PSEAE	O33877 pseudomonas

ALIGNMENTS

RESULT 1
ACVS EMENI
ID ACVS EMENI STANDARD; PRT; 3770 AA.
AC P27742;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteiny-D-valine
DE synthetase) (ACV synthetase) (ACVS).
GN ACVA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=G191;
RX MEDLINE=91286299; PubMed=2061333;
RA Maccabe A.P., van Liempt H., Pallissa H., Unkles S.E., Riach M.B.R., Pfeifer E., von Doehren H., Kinghorn J.R.;
RT "Delta-(L-alpha-aminoadipyl)-L-cysteiny-D-valine synthetase from Aspergillus nidulans. Molecular characterization of the acva gene encoding the first enzyme of the penicillin biosynthetic pathway."; J. Biol. Chem. 266:12646-12654(1991).
CC -!- FUNCTION: Each of the constituent amino acids of the tripeptide acv are activated as aminoacyl-adenylates with peptide bonds formed through the participation of amino acid thiolester intermediates.
CC -!- CATALYTIC ACTIVITY: L-2-aminohexanedioate + L-cysteine + L-valine + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteiny-D-valine + 3 AMP + 3 diphosphate.
CC -!- COFACTOR: Contains 3 covalently bound phosphopantetheines (Potential).
CC -!- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme family.
CC -!- SIMILARITY: Contains 3 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X54853; CAA38631.1; --
CC PTR: A40889; A40889.
CC HSP; P14687; IAMU.
CC InterPro: IPR000873; AMP-bind.
CC InterPro: IPR001242; Condensatn.
CC InterPro: IPR006163; Pp bind.
CC InterPro: IPR006162; Ppantne S.
CC InterPro: IPR000379; Ser_estfs.

```

DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00668; Condensation; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00455; AMP BINDING; 3.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
DR Repeat; Antibiotic biosynthesis; Multifunctional enzyme;
KW Repeat; Phosphopantetheine.
FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 850 919 ACYL CARRIER (ACP) 1.
FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 3.
FT BINDING 882 882 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 1965 1965 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3050 3050 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACT SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3770 AA; 422448 MW; CB66B6D232A58CB0 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 3770;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 SVIAQWYKVMGSGT 16
Db 468 SLTSKQLAYVTYSGT 483

RESULT 2
ID SRF2_BACSU STANDARD; PRT; 3587 AA.
AC Q04747;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Surfactin synthetase subunit 2.
GN SRFAB OR SRF2 OR COML OR BSU03490.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_taxID=1423;
EN [1]
RP SEQUENCE OF 1-3077 FROM N.A.
RC STRAIN=168;
RX MEDLINE=93181186; PubMed=8441623;
RA Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA Zuber P., Yamane K.;
RT "Nucleotide sequence of 5' portion of srfA that contains the region
RT required for competence establishment in Bacillus subtilis.";
RL Nucleic Acids Res. 21:93-97(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=93360813; PubMed=8355609;
RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA Venema G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis.";
RL Mol. Microbiol. 8:821-831(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes.";
RL Microbiology 142:3047-3056(1996).
RN [4]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerth J.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi H., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Halbert G., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidine A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudenga B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognotti A.,
RA Totsato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 514-800 FROM N.A.
RC STRAIN=ATCC 2132; PubMed=1601288;
RX MEDLINE=92290255; PubMed=1601288;
RA Borchert S., Patil S.S., Marahiel M.A.;
RT "Identification of putative multifunctional peptide synthetase genes
RT using highly conserved oligonucleotide sequences derived from known
RT synthetases.";
RL FEMS Microbiol. Lett. 71:175-180(1992).
CC -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC -!- COFACTOR: Contains 3 covalently bound phosphopantetheinates.
CC -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC -!- SIMILARITY: Belongs to the Amp-dependent AMP-binding enzyme
CC family.
CC -!- SIMILARITY: Contains 3 acyl carrier domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13262; BAA02523.1; -
CC EMBL; X70356; CAA49817.1; -
CC EMBL; D50453; BAA08983.1; -
CC EMBL; Z9105; CAB12143.1; -
CC EMBL; X65835; CAA46678.1; -
CC PIR; I40486; I40486.
CC HSSP; P14687; 1AMU.
CC Subtilist; BG10169; srfAB.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensatn.
CC InterPro; IPR006163; Pp bind.
CC InterPro; IPR006162; Ppantne S.
CC Pfam; PF00501; AMP-binding; 3.
CC Pfam; PF00668; Condensation; 4.

```

DR Pfam; PF00550; pp-binding; 3.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
DR PROSITE; PS00455; AMP BINDING; 3.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
KW Ligase; Anticibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW Multifunctional enzyme; Repeat; Complete proteome.
FT REPEAT ? 1040 DOMAIN 1 (VAL-ACTIVATING).
FT REPEAT ? 2096 DOMAIN 2 (ASP-ACTIVATING).
FT REPEAT ? 3114 DOMAIN 3 (D-LEU-ACTIVATING).
FT DOMAIN 970 ACYL CARRIER (ACP) 1.
FT DOMAIN 1036 ACYL CARRIER (ACP) 2.
FT DOMAIN 2015 ACYL CARRIER (ACP) 3.
FT DOMAIN 3043 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 999 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 2045 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 3073 PHOSPHOPANTHETHEINE (POTENTIAL).
FT CONFLICT 33 S -> F (IN REF. 1).
FT CONFLICT 42 A -> G (IN REF. 1).
FT CONFLICT 110 Q -> D (IN REF. 1).
FT CONFLICT 113 ROA -> AQG (IN REF. 1).
FT CONFLICT 139 A -> V (IN REF. 1).
FT CONFLICT 259 L -> W (IN REF. 1).
FT CONFLICT 309 R -> A (IN REF. 1).
FT CONFLICT 478 TPA -> SRP (IN REF. 1).
FT CONFLICT 596 MISSING (IN REF. 5).
FT CONFLICT 648 A -> R (IN REF. 1).
FT CONFLICT 680 RHV -> ETL (IN REF. 1).
FT CONFLICT 694 BOSIT -> DKRIS (IN REF. 5).
FT CONFLICT 788 M -> L (IN REF. 5).
FT CONFLICT 939 PL -> LV (IN REF. 1).
FT CONFLICT 1038 N -> I (IN REF. 1).
FT CONFLICT 1133 H -> Q (IN REF. 1).
FT CONFLICT 1310 V -> C (IN REF. 1).
FT CONFLICT 1333 G -> V (IN REF. 1).
FT CONFLICT 1384 P -> R (IN REF. 1).
FT CONFLICT 1582 G -> E (IN REF. 1).
FT CONFLICT 1677 KRRADG -> E (IN REF. 1).
FT CONFLICT 1700 C -> S (IN REF. 1).
FT CONFLICT 1755 F -> K (IN REF. 1).
FT CONFLICT 1787 T -> S (IN REF. 1).
FT CONFLICT 1801 GAIAGRVLDYEDFAKRPPTG -> APSPGGLICMSRCIC
FT CONFLICT 1915 ETPDNR (IN REF. 1).
FT CONFLICT 1916 LG -> PK (IN REF. 1).
FT CONFLICT 2075 R -> C (IN REF. 1).
FT CONFLICT 2079 A -> V (IN REF. 1).
FT CONFLICT 2141 ARLTPT -> LRDSIN (IN REF. 1).
FT CONFLICT 2445 E -> Q (IN REF. 1).
FT CONFLICT 2485 ATDLF -> RQICS (IN REF. 1).
FT CONFLICT 2566 TVHQLFEETVQRHKORPAVTV -> DGCISYSKRLSSATKT
FT CONFLICT 2608 ARLSHT (IN REF. 1).
FT CONFLICT 2645 MSAAVLGV -> KCPPRCSAS (IN REF. 1).
FT CONFLICT 2713 KL -> NV (IN REF. 1).
FT CONFLICT 2723 D -> H (IN REF. 1).
FT CONFLICT 2876 D -> H (IN REF. 1).
FT CONFLICT 2899 GELCVA -> RALRG (IN REF. 1).
FT CONFLICT 2900 RF -> L (IN REF. 1).
FT CONFLICT 2958 EDR -> QDA (IN REF. 1).
FT CONFLICT 2964 R -> A (IN REF. 1).
SQ SEQUENCE 3587 AA; 401248 MW; 6B0B05A9F32054D CRC64;

Query Match 49.4%; Score 41; DB 1; Length 3587;
Best Local Similarity 43.8%; Pred. No. 62;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Qy 1 SVIAKQMTYKVMGSGT 16
Db 2687 AVTAENLAMYITSGT 2702

RESULT 3
ID_SF1_BACSU STANDARD; PRT; 3588 AA.
AC P27206;

DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Surfactin synthetase subunit 1.
GN SRFAA OR SRFAI OR SRFA OR BSU03480.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=1423;
[1] SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93181186; PubMed=8441623;
RA Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
Zuber P., Yamane K.;
RT "Nucleotide sequence of 5' portion of srfA that contains the region
required for competence establishment in Bacillus subtilis";
RL Nucleic Acids Res. 21:93-97(1993).
[2] SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=93360813; PubMed=8355609;
RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
Venema G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
synthesis in Bacillus subtilis";
RL Mol. Microbiol. 8:821-831(1993).
[3] SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
chromosome: determination of the sequence of a 146 kb segment and
identification of 113 genes";
RL Microbiology 142:3047-3056(1996).
[4] SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Paro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,
Presecan B., Pujic C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis";
RL Nature 390:249-256(1997).
[5] SEQUENCE OF 1-460 FROM N.A.
RX MEDLINE=91154134; PubMed=1847909;


```

CC CC      -!- SIMILARITY: Contains 1 engC GTPase domain.
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: AP001515; BAB06222.1; -.
CC CC      FIRM: G83962; G83962.
CC CC      HAMAP: MF_01820; -. 1.
CC CC      InterPro: IPR004881; DUF258.
CC CC      Pfam: PF03193; DUF258; 1.
CC CC      TIGRFAMs: TIGR00157; TIGR00157; 1.
CC CC      PROSITE: PS50936; ENG_C_GTPASE; 1.
CC CC      KW      Hydrolase; GTP-binding; Complete proteome.
CC CC      FT      DOMAIN 72 221      ENG_C_GTPASE.
CC CC      FT      NP_BIND 112 115      GTP (PROBABLE).
CC CC      FT      NP_BIND 166 173      GTP (PROBABLE).
CC CC      FT      NP_BIND 216 220      GTP (PROBABLE).
CC CC      FT      SITE 247 260      KNUCKLE-LIKE CYSTEINE CLUSTER.
CC CC      SQ      SEQUENCE 294 AA; 33449 MW; 7DC80C97AFA93A9B CRC64;
CC CC      Query Match 48.2%; Score 40; DB 1; Length 294;
CC CC      Best Local Similarity 46.2%; Pred. No. 8.2;
CC CC      Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
CC CC      QY      5 KQMTYKYVMSGTV 17
CC CC      DB      133 EQLGYKYLSTI 145
CC CC      RESULT 5
CC CC      SN2A_SCHGR
CC CC      ID      SN2A_SCHGR STANDARD; PRT; 697 AA.
CC CC      AC      Q9XZC8;
CC CC      DT      30-MAY-2000 (Rel. 39, Created)
CC CC      DT      10-MAY-2000 (Rel. 39, Last sequence update)
CC CC      DT      10-OCT-2003 (Rel. 42, Last annotation update)
CC CC      DE      Semaphorin 2A precursor (Sema 2A) (Sema II).
CC CC      GN      Sema-2A.
CC CC      OS      Schistocerca gregaria (Desert locust).
CC CC      OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
CC CC      OC      Neoptera; Orthopteroidea; Orthoptera; Gaelelfera; Acridoidea;
CC CC      OC      Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.
CC CC      OX      NCBI_TaxID=7010;
CC CC      RN      [1]
CC CC      RP      SEQUENCE FROM N.A.
CC CC      RC      TISSUE=Embryo;
CC CC      RX      MEDLINE=99203450; PubMed=10101134;
CC CC      RA      Isbister C.M., Tsai A., Wong S.T., Kolodkin A.L., O'Connor T.P.;
CC CC      RT      "Discrete roles for secreted and transmembrane semaphorins in neuronal
CC CC      RL      growth cone guidance in vivo.";
CC CC      CC      Development 126:2007-2019(1999).
CC CC      -!- FUNCTION: ACTS AS CHEMOREPULSIVE GUIDANCE MOLECULE CRITICAL FOR
CC CC      AND SUBSEQUENT PATHFINDING EVENTS OF THE TI AXON PROJECTION.
CC CC      -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC CC      -!- TISSUE SPECIFICITY: EXPRESSED IN A GRADIENT IN THE DEVELOPING LIMB
CC CC      BUD EPITHELIUM DURING TI PIONEER AXON OUTGROWTH.
CC CC      -!- SIMILARITY: Belongs to the semaphorin family.
CC CC      -!- SIMILARITY: Contains 1 Sema domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).

```

```

CC CC      EMBL: AF134904; AAD30114.1; -.
CC CC      InterPro: IPR007110; IG-like.
CC CC      InterPro: IPR003599; IG.
CC CC      InterPro: IPR001627; Sema.
CC CC      Pfam: PF01403; Sema; 1.
CC CC      SMART: SM00409; IG; 1.
CC CC      SMART: SM00630; Sema; 1.
CC CC      PROSITE: PS50835; IG_LIKE; 1.
CC CC      KW      Signal; Immunoglobulin domain; Neurogenesis; Developmental protein;
CC CC      Glycoprotein.
CC CC      FT      SIGNAL 1 20      POTENTIAL.
CC CC      FT      CHAIN 21 697      SEMAPHORIN 2A.
CC CC      FT      DOMAIN 231 515      Sema.
CC CC      FT      DOMAIN 526 634      IG-LIKE C2-TYPE.
CC CC      FT      DISULFID 618 654      BY SIMILARITY.
CC CC      FT      CARBOHYD 63 63      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 66 66      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 132 132      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 198 198      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 283 283      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 369 369      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 534 534      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 629 629      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      FT      CARBOHYD 679 679      N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC      SQ      SEQUENCE 697 AA; 78673 MW; 7FE55AB4A965E1E4 CRC64;
CC CC      Query Match 48.2%; Score 40; DB 1; Length 697;
CC CC      Best Local Similarity 43.8%; Pred. No. 19;
CC CC      Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
CC CC      QY      1 SVIAKQMTYKYVMSGT 16
CC CC      DB      419 NLVGKNMEYIYYAGT 434
CC CC      RESULT 6
CC CC      OL7A_MOUSE
CC CC      ID      OL7A_MOUSE STANDARD; PRT; 161 AA.
CC CC      AC      P34985;
CC CC      DT      01-FEB-1994 (Rel. 28, Created)
CC CC      DT      01-FEB-1994 (Rel. 28, Last sequence update)
CC CC      DT      15-JUL-1998 (Rel. 36, Last annotation update)
CC CC      DE      Olfactory receptor 7A (K18) (Fragment).
CC CC      GN      OLFR7.
CC CC      OS      Mus musculus (Mouse).
CC CC      OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC CC      OX      NCBI_TaxID=10090;
CC CC      RN      [1]
CC CC      RP      SEQUENCE FROM N.A.
CC CC      RC      STRAIN=C57BL/6J; TISSUE=Liver;
CC CC      RX      MEDLINE=93258822; PubMed=7683976;
CC CC      RA      Ressler K.J., Sullivan S.L., Buck L.B.;
CC CC      RT      "A zonal organization of odorant receptor gene expression in the
CC CC      RL      olfactory epithelium.";
CC CC      CC      Cell 73:597-609(1993).
CC CC      -!- FUNCTION: Putative odorant receptor.
CC CC      -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC      -!- TISSUE SPECIFICITY: Olfactory epithelium.
CC CC      -!- MISCELLANEOUS: OLFACTORY RECEPTORS 7A-7I ARE PRODUCED BY EIGHT
CC CC      DIFFERENT GENES WITHIN THE OLFR7 COMPLEX.
CC CC      -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      EMBL: L14568; AAA39852.1; -.

```

DR PIR; A40745; A40745.
DR MGD; MGI:104712; Olfr7.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; PARTIAL.
DR PROSITE; PS00262; G PROTEIN RECEPTOR FL 2; 1.
DR G-protein coupled receptor; Transmembrane; Multigene family;
KW Olfaction.
FT NON_TER 1 18
FT DOMAIN <1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 37 4 (POTENTIAL).
FT DOMAIN 38 75 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 76 98 5 (POTENTIAL).
FT DOMAIN 99 115 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 116 139 6 (POTENTIAL).
FT DOMAIN 140 151 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 152 >161 7 (POTENTIAL).
FT NON_TER 161 161
SQ SEQUENCE 161 AA; 17562 MW; 7A5140BB1EFB7FB7 CRC64;

Query Match 47.0%; Score 39; DB 1; Length 161;
Best Local Similarity 66.7%; Pred. No. 7;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 IAKQMTYKVMYS 14
| : : : : :
DB 5 ICKPLTYKVMIS 16
| : : : : :

RESULT 7
ID CBP4_YEAST STANDARD; PRT; 170 AA.
AC P37267;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CBP4 protein, mitochondrial precursor.
GN CBP4 OR YGR174C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94342301; PubMed=8063753;
RA Crivellone M.D.;
RT "Characterization of CBP4, a new gene essential for the expression of
ubiquinol-cytochrome c reductase in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 269:21284-21292(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Hebling U., Hofmann B., Delius H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Essential for the assembly and/or stability of
ubiquinol-cytochrome c reductase. It has a direct effect on the
correct occurrence of the Rieske protein, core 4, core 5 and
apocytochrome b; it may either be involved in post-translational
modification of the subunits or in the assembly of the enzyme.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL. ASSOCIATED WITH THE INNER
MEMBRANE.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U10700; AAA61566.1; -
EMBL; Z72959; CAA97200.1; -
PIR; S64488; S64488.
GermOnline; 141486; -.

DR SGD; S0003406; CBP4.
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 170 CBP4 PROTEIN.
FT CONFLICT 65 65 S -> F (IN REF. 1).
SQ SEQUENCE 170 AA; 20219 MW; D88F92EADF0B366E CRC64;

Query Match 47.0%; Score 39; DB 1; Length 170;
Best Local Similarity 61.5%; Pred. No. 7.3;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKVMY 13
| : : : : :
DB 10 AVIAKQRYKHYL 22
| : : : : :

RESULT 8
ID PLSX_AQUAE STANDARD; PRT; 337 AA.
AC O67186;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fatty acid/phospholipid synthesis protein plsx.
GN PLSX OR AQ_1101.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RC MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujaay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: Not known, probably involved in fatty acid or
phospholipid synthesis (By similarity).
CC -!- SIMILARITY: Belongs to the plsx family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE000723; AAC07145.1; -
PIR; G70394; G70394.
DR HAMAP; MF_00019; -; 1.
DR InterPro; IPR003664; FA_synthesis.
DR Pfam; PF02504; FA_synthesis; 1.
DR ProDom; PD006974; FA_synthesis; 1.
DR TIGRFAMs; TIGR00182; plsx; 1.
KW Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.
SQ SEQUENCE 337 AA; 36266 MW; C6E51574FA15D508 CRC64;

Query Match 47.0%; Score 39; DB 1; Length 337;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 AKQMTYKVMYSG 15
| : : : : :
DB 26 AKELGYKYLNG 37
| : : : : :

RESULT 9
SERC_NEIMA
ID SERC_NEIMA STANDARD; PRT; 368 AA.
AC O34370; O33382; O33383; O33384; O33386;

FT BINDING 203 203 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 368 AA; 41393 MW; 97DFCE52BBE5E021 CRC64;

Query Match 47.0%; Score 39; DB 1; Length 368;
Best Local Similarity 70.0%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKYVMSGTV 17
||:|||||
Db 248 TYAIYMSGVL 257

RESULT 11
ID SP98_YEAST STANDARD; PRT; 846 AA.
AC P33540;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Spindle pole body component SPC98.
GN SPC98 OR YNL126W OR N1222 OR N1879.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96109932; PubMed=8619318;
RA Mallet L., Bussereau F., Jacquet M.;
RA "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT MRP2, CAP/SRV2, NAM9, FKBI/FPRI/RP1, MOM22 and CPT1, predicts an
RT adenosine deaminase gene and 14 new open reading frames.";
RL Yeast 11:1195-1209(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97245296; PubMed=9090055;
RA de Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA Pallavicini A., Lanfranchi G., Valle G.;
RA "The DNA sequence of cosmid 14-13b from chromosome XIV of
RT Saccharomyces cerevisiae reveals an unusually high number of
RT overlapping open reading frames.";
RL Yeast 13:261-266(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96324398; PubMed=8670895;
RA Geissler S., Pereira G., Spang A., Knop M., Soues S., Kilmartin J.V.,
RA Schiebel E.;
RA "The spindle pole body component SPC98p interacts with the
RT gamma-tubulin-like Tub4p of Saccharomyces cerevisiae at the sites of
RT microtubule attachment.";
RL EMBO J. 15:3899-3911(1996).
CC -!- FUNCTION: Involved in microtubule organization by the microtubule
CC organizing centre, the spindle pole body (SPB). Probably part of
CC the microtubule attachment site at the SPB.
CC -!- SUBUNIT: Interacts with TUB4 and SPC97.
CC -!- SUPPLEMENTARY INFORMATION: Nuclear.
CC -!- SIMILARITY: Belongs to the GCP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 246843; CAA86899.1; -
CC EMBL; 269382; CAA93378.1; -
CC EMBL; 271402; CAA96007.1; -
CC FIR; S59262; S59262.
CC GeneOnline; 143132; -
CC SGD; S0005070; SPC98.
DR

DR GO; GO:0005822; C:inner plaque of spindle pole body; IDA.
DR GO; GO:0005824; C:outer plaque of spindle pole body; IDA.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; IPI.
DR GO; GO:0007020; P:microtubule nucleation; IPI.
DR GO; GO:0000071; P:mitotic spindle assembly (sensu Saccharomycetes); IMP.
DR InterPro; IPR007259; SPC97_Spc98.
DR Pfam; PF04130; SPC97_Spc98; 1.
KW Microtubule; Nuclear protein.
SQ SEQUENCE 846 AA; 98226 MW; 803048B05D5E5105 CRC64;

Query Match 47.0%; Score 39; DB 1; Length 846;
Best Local Similarity 46.2%; Pred. No. 35;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 IAKOMTYKYVMSG 15
||:|||||
Db 394 IPKELAYKIPWIG 406

RESULT 12
ID ACVS_CEPAC STANDARD; PRT; 3712 AA.
AC P25464;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteiny-D-valine
DE synthetase) (ACV synthetase) (ACVS).
GN PCBAB.
OS Cephalosporium acremonium (Acremonium chrysogenum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Mitosporic Hypocreaceae;
OC Acremonium.
OX NCBI_TaxID=5044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91177827; PubMed=1706706;
RA Gutierrez S., Diez B., Montenegro E., Martin J.F.;
RA "Characterization of the Cephalosporium acremonium pcbAB gene
RT encoding alpha-aminoadipyl-L-cysteiny-D-valine synthetase, a large
RT multidomain peptide synthetase: linkage to the pcbC gene as a cluster
RT of early cephalosporin biosynthetic genes and evidence of multiple
RT functional domains.";
RL J. Bacteriol. 173:2354-2365(1991).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 11550; PubMed=2076552;
RX MEDLINE=91168300; PubMed=2076552;
RA Hoskins J.A., O'Callaghan N., Queener S.W., Cantwell C.A., Wood J.S.,
RA Chen V.J., Skatrud P.L.;
RA "Gene disruption of the pcbAB gene encoding ACV synthetase in
RT Cephalosporium acremonium.";
RL Curr. Genet. 18:523-530(1990).
CC -!- FUNCTION: Each of the constituent amino acids of the tripeptide
CC acv are activated as aminoacyl-adenylates with peptide bonds
CC formed through the participation of amino acid thioester
CC intermediates.
CC -!- CATALYTIC ACTIVITY: L-2-amino-6-oxoheptanoate + L-cysteine + L-valine
CC + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteiny-D-valine +
CC 3 AMP + 3 diphosphate.
CC -!- COFACTOR: Contains 3 covalently bound phosphopantetheines
CC (Potential).
CC -!- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -!- SIMILARITY: Contains 3 acyl carrier domains.
CC FIR; A38531; YGCEVC.
DR HSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR006163; Pp-bind.
DR InterPro; IPR006162; Ppantne_S.
DR

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; S67816; AAB28883.2; ..
 DR EMBL; AE000149; AAC73537.1; ALT INIT.
 DR EMBL; U82664; AAB40190.1; ALT INIT.
 DR EMBL; AE016756; AAN79024.1; ALT_INIT.
 DR EcoGene; EG2182; yaJG.
 DR InterPro; IPR005619; Lipoprotein 16.
 DR InterPro; IPR000437; Prok lipoprot S.
 DR Pfam; PF03923; Lipoprotein 16; 1.
 DR ProDom; PD036382; lipoprotein 16; 1.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 DR Hypothetical protein; Membrane; Lipoprotein; Signal;
 KW Complete proteome; Palmitate.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 192 HYPOTHETICAL LIPOPROTEIN YAJG.
 FT LIPID 18 18 N-palmitoyl cysteine (Potential).
 FT LIPID 18 18 S-diacylglycerol cysteine (Potential).
 SQ SEQUENCE 192 AA; 20950 MW; 9B9E6568E9253451 CRC64;

Query Match 45.8%; Score 38; DB 1; Length 192;
 Best Local Similarity 57.1%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 VIAKQMTYKVMVG 15
 Db 83 VLEKQMTARGVMVG 96

RESULT 15
 BChL RHOC
 ID BChL RHOC STANDARD; PRT; 304 AA.
 AC P26237;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Light-independent protochlorophyllide reductase iron-sulfur ATP-
 DE binding protein (EC 1.18.-.-) (LI-POR subunit L) (DPOR subunit L).
 GN BChL.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Rhodospirillaceae; Rhodospirillum.
 OX NCBI_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368552; PubMed=2203738;
 RA Yang Z., Bauer C.E.;
 RT "Rhodospirillum rubrum genes involved in early steps of the
 RT bacteriochlorophyll biosynthetic pathway";
 RL J. Bacteriol. 172:5001-5010 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003 / St Louis;
 RA Burke D.H., Alberti M., Armstrong G.A., Hearst J.E.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=84259352; PubMed=6744416;
 RA Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
 RT "Nucleotide and deduced polypeptide sequences of the photosynthetic
 RT reaction-center, B870 antenna, and flanking polypeptides from R.
 RT capsulata";
 RL Cell 37:949-957 (1984).
 RN [4]
 RN CHARACTERIZATION.
 RC STRAIN=SB1003 / CB1029;
 RX MEDLINE=20378986; PubMed=10811655;
 RA Fujita Y., Bauer C.E.;
 RT "Reconstitution of light-independent protochlorophyllide reductase

RT from purified bchl and bchl-bchl subunits. In vitro confirmation of
 RT nitrogenase-like features of a bacteriochlorophyll biosynthesis
 RT enzyme";
 RL J. Biol. Chem. 275:23583-23588 (2000).
 RN [5]
 RP CHARACTERIZATION.

RA Fujita Y.;
 RL Unpublished observations (JUL-2001).
 CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
 CC protochlorophyllide (pchl) to form chlorophyllide a (Chl).
 CC This reaction is light-independent.
 CC -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
 CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
 CC of three subunits; bchl, bchl and bchl. Homodimer of bchl subunit
 CC (By similarity).
 CC -!- SIMILARITY: Belongs to the nifH / bchl / chl family.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M34843; AAA26098.1; ..
 DR EMBL; Z11165; CAA77523.1; ..
 DR EMBL; K01183; -; NOT_ANNOTATED_CDS.
 DR PIR; B36716; B36716.
 DR PIR; H28771; H28771.
 DR HSSP; P00456; 1CP2.
 DR HAMAP; MF_00355; -; 1.
 DR InterPro; IPR000392; NitrogenaseII.
 DR InterPro; IPR005971; Protochl_reductF.
 DR Pfam; PF00142; fer4_nifh; 1.
 DR PRINTS; PR00091; NITROGNASEII.
 DR TIGRfam; TIGR01281; DPOR_bchl; 1.
 DR PROSITE; PS00746; NIFH_FRXC 1; 1.
 DR PROSITE; PS00692; NIFH_FRXC 2; 1.
 KW Oxidoreductase; Photosynthesis; Bacteriochlorophyll biosynthesis;
 KW ATP-binding; Iron-sulfur; 4Fe-4S.

FT NP_BIND 43 50 ATP (POTENTIAL).
 FT METAL 131 131 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 165 165 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 304 AA; 33204 MW; 3A49C39BCF15AEC64;
 Query Match 45.8%; Score 38; DB 1; Length 304;
 Best Local Similarity 47.1%; Pred. No. 20;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVMVG 17
 Db 198 AVQAKSVNKKVRLAGCV 214

Search completed: August 12, 2004, 06:20:06
 Job time : 3.04508 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 4.42418 Seconds
(without alignments)
1212.395 Million cell updates/sec

Title: US-09-890-463-2

Perfect score: 83

Sequence: 1 SVIAKQMTYKYVMSGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_podent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	83	100.0	221	5	Q95P04
2	45	54.2	398	16	Q97LU2
3	43	51.8	389	16	Q8XIP1
4	43	51.8	424	4	Q9UF27
5	43	51.8	4243	4	Q86W11
6	41	49.4	227	5	Q8MU45
7	41	49.4	227	5	Q95W11
8	41	49.4	227	5	Q95W86
9	41	49.4	592	17	Q96X15
10	41	49.4	3583	2	Q45675
11	40	48.2	218	16	Q9C165
12	40	48.2	225	5	Q9U678
13	40	48.2	230	5	Q3GTJ7
14	40	48.2	263	16	Q97FT4
15	40	48.2	268	10	Q7XDZ1
16	40	48.2	285	16	Q9KBK0

ALIGNMENTS

RESULT 1

Q95P04 PRELIMINARY; PRT; 221 AA.
ID Q95P04;
AC Q95P04;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Gonipora tenuidens.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Fungiina; Poritidae; Gonipora.
OX NCBI_TaxID=75301;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Pradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF383156; AAL27542.1;
DR GO; GO:0006031; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 24918 MW; 93F9F4B5C2003CB4 CRC64;

Query Match 100.0%; Score 83; DB 5; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKYVMSGTV 17

Dd 2 SVIAKQMTYKYVMSGTV 18

RESULT 2

Q97LU2

```

ID Q97LU2 PRELIMINARY; PRT; 398 AA.
AC Q97LU2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protein of short-chain alcohol dehydrogenase family.
GN CAC0462.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838 (2001).
DR EMBL; AB007561; AAK78442.1; -.
DR PIR; G96956; G96956.
KW Complete proteome.
SQ SEQUENCE 398 AA; 45650 MW; 59324A21CA466DFC CRC64;

Query Match 54.2%; Score 45; DB 16; Length 398;
Best Local Similarity 47.1%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKYVMSGTV 17
DB 224 SYIGSPRTYKYREGTI 240

RESULT 3
Q8XIP1 PRELIMINARY; PRT; 389 AA.
AC Q8XIP1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein CPE2074.
GN CPE2074.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Onshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
DR EMBL; AF003192; BAB81780.1; -.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 389 AA; 43138 MW; 36E1230CC803E7C5 CRC64;

Query Match 51.8%; Score 43; DB 16; Length 389;
Best Local Similarity 41.2%; Pred. No. 32;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKYVMSGTV 17
DB 223 SYIGPVTYPIYREGTI 239

RESULT 4
Q9UF27 PRELIMINARY; PRT; 424 AA.
AC Q9UF27;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP586C1021.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Uterus;
RX Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AL133640; CAB63761.1; -.
DR PIR; T43498; T43498.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 424 AA; 46402 MW; 35523FD7C62313A2 CRC64;

Query Match 51.8%; Score 43; DB 4; Length 424;
Best Local Similarity 43.8%; Pred. No. 34;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKYVMSGT 16
DB 1 SIVALKSYEYFTGT 16

RESULT 5
Q86W11 PRELIMINARY; PRT; 4243 AA.
ID Q86W11;
AC Q86W11;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fibrocystin L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22508206; PubMed=12620974;
RA Hogan M.C., Griffin M.D., Rossetti S., Torres V.E., Ward C.J.,
RA Harris P.C.;
RT "PKHDLL1, a homolog of the autosomal recessive polycystic kidney
RT disease gene, encodes a receptor with inducible T lymphocyte
RT expression.";
RL Hum. Mol. Genet. 12:685-698 (2003).
DR EMBL; AY219181; AAO60072.1; -.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR002909; IPT_TIG.
DR InterPro; IPR006626; Psh1.
DR Pfam; PF01833; TIG; 14.
DR SMART; SM00429; IPT; 14.
DR SMART; SM00710; Psh1; 10.
SQ SEQUENCE 4243 AA; 465745 MW; 36FE9DB63F4931E7 CRC64;

Query Match 51.8%; Score 43; DB 4; Length 4243;
Best Local Similarity 43.8%; Pred. No. 3.6e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKYVMSGT 16
DB 3820 SIVALKSYEYFTGT 3835

RESULT 6
Q8MU45 PRELIMINARY; PRT; 227 AA.
ID Q8MU45

```

```
AC Q8MU45;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein-like protein.
OS Condylactis gigantea (Giant anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynantheae; Actiniidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RA Matz M.V., Lukyanov S.A.;
RT "Diversity and evolution of GFP-like fluorescent proteins.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037777; AAK71343.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25384 MW; D3C6B02F490F3D21 CRC64;

Query Match          49.4%; Score 41; DB 5; Length 227;
Best Local Similarity 43.8%; Pred. No. 41;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIAQMTYKVMYSGTV 17
   : : | : | | | |
Db 4 LLKESMRKIYMEGV 19

RESULT 7
ID Q95W11 PRELIMINARY; PRT; 227 AA.
AC Q95W11;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Condylactis passiflora.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynantheae; Actiniidae; Condylactis.
OX NCBI_TaxID=175772;
RN [1]
RP SEQUENCE FROM N.A.
RA Matz M.V., Lukyanov S.A.;
RT "Diversity and evolution of GFP-like fluorescent proteins.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037777; AAK71343.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25384 MW; D3C6B02F490F3D21 CRC64;

Query Match          49.4%; Score 41; DB 5; Length 227;
Best Local Similarity 43.8%; Pred. No. 41;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIAQMTYKVMYSGTV 17
   : : | : | | | |
Db 4 LLKESMRKIYMEGV 19

RESULT 8
ID Q95W86 PRELIMINARY; PRT; 227 AA.
AC Q95W86;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Condylactis gigantea (Giant anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynantheae; Actiniidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RA Gurekaya N.G., Fradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF363775; AAL27541.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25416 MW; BCFASA4CBCE1B3F7 CRC64;

Query Match          49.4%; Score 41; DB 5; Length 227;
Best Local Similarity 43.8%; Pred. No. 41;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIAQMTYKVMYSGTV 17
   : : | : | | | |
Db 4 LLKESMRKIYMEGV 19

RESULT 9
ID Q96YI5 PRELIMINARY; PRT; 592 AA.
AC Q96YI5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative glucosamine--fructose-6-phosphate aminotransferase.
GN ST2186.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RA Kwarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000989; BAB67292.1; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004360; F:glutamine-fructose-6-phosphate transaminase. . .; IEA.
DR GO; GO:0005529; F:sugar binding; IEA.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR00583; GATase_2.
DR InterPro; IPR005855; Glms.
DR InterPro; IPR01347; SIS.
DR Pfam; PF00310; GATase_2; 1.
DR Pfam; PF01380; SIS; 2.
DR TIGRFAMs; TIGR01135; glms; 1.
DR PROSITE; PS00443; GATASE TYPE II; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 592 AA; 65796 MW; 3CED613D9A0EB7ED CRC64;
```


OC Discosomatidae; Discosoma.
 RN NCBI_TaxID=86600;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99436614; PubMed=10504696;
 RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
 RA Markelov M.L., Lukyanov S.A.;
 RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
 RL Nat. Biotechnol. 17:969-973(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
 RA Markelov M.L., Lukyanov S.A.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF168419; AAF03369.1; -.
 DR PDB; 1G7K; 07-NOV-01.
 DR PDB; 1GKX; 06-DEC-00.
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP_like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP/NORESCENT.
 DR ProDom; PD013756; Green.fl protein; 1.
 SQ SEQUENCE 225 AA; 25931 MW; FBF9A5369778F689 CRC64;

Query Match 48.2%; Score 40; DB 5; Length 225;
 Best Local Similarity 52.9%; Pred. No. 62;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 SVIAKQMTYKVMSTGTV 17

Db 6 NVIKEFMRFKVRMEGTV 22

RESULT 13

O9GTJ7
 ID O9GTJ7 PRELIMINARY; PRT; 230 AA.
 AC O9GTJ7;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Red fluorescent protein.
 GN FP593.
 OS Discosoma sp. SSAL-2000.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
 OC Discosomatidae; Discosoma.
 OX NCBI_TaxID=137428;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20434599; PubMed=10981720;
 RA Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
 RA Lukyanov S.A.;
 RT "Novel fluorescent protein from Discosoma coral and its mutants
 RT possesses a unique far-red fluorescence.";
 RL FEBS Lett. 479:127-130(2000).
 DR EMBL; AF272711; AAG16224.1; -.
 DR HSP; P42212; 1BFP.
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR InterPro; IPR000786; Green.fl protein.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP/NORESCENT.
 DR ProDom; PD013756; Green.fl protein; 1.
 SQ SEQUENCE 230 AA; 26370 MW; 5215B1B436D67E51 CRC64;

Query Match 48.2%; Score 40; DB 5; Length 230;
 Best Local Similarity 52.9%; Pred. No. 63;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 SVIAKQMTYKVMSTGTV 17

Db 6 NVIKEFMRFKVRMEGTV 22

RESULT 14

O97FT4
 ID O97FT4 PRELIMINARY; PRT; 263 AA.
 AC O97FT4;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Hypothetical protein CAC2643.
 GN CAC2643.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AE007761; AAK80590.1; -.
 DR PIR; C37225; C97225.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 263 AA; 29220 MW; 41D648F7237A42AB CRC64;

Query Match 48.2%; Score 40; DB 16; Length 263;
 Best Local Similarity 43.8%; Pred. No. 72;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 SVIAKQMTYKVMSTGTV 16

Db 152 SGVQDITYKAYTSGS 167

RESULT 15

O7XDZ1
 ID O7XDZ1 PRELIMINARY; PRT; 268 AA.
 AC O7XDZ1;
 DT 01-OCT-2003 (TReMBLrel. 25, Created)
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Putative AP2-domain DNA-binding protein.
 GN OSJNBA0094K20.6.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartodeae; Oryzaceae; Oryza.
 OX NCBI_TaxID=33947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Nipponbare;
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice
 RT chromosome 10.";
 RL science 300:1566-1569(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Nipponbare;
 RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE017099; AAP54000.1; -.
 KW DNA-binding.
 SQ SEQUENCE 268 AA; 30728 MW; 58DC0E0881BEA6E1 CRC64;

Query Match 48.2%; Score 40; DB 10; Length 268;
 Best Local Similarity 57.1%; Pred. No. 74;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 IAKOMTYKVMSGT 16
Db 94 LLKIMYKVYADGT 107

Search completed: August 12, 2004, 06:19:35
Job time : 7.67418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:19:43 ; Search time 35.7766 Seconds
(without alignments)
149.169 Million cell updates/sec

Title: US-09-890-463-2

Perfect score: 83

Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	51.8	107	10	US-09-994-595-43
2	42	50.6	47	12	US-10-424-599-172498
3	41	49.4	225	14	US-10-315-920-6
4	41	49.4	225	15	US-10-442-148A-7
5	41	49.4	239	15	US-10-442-148A-8
6	40	48.2	26	14	US-10-081-864-25
7	40	48.2	205	13	US-10-006-922-46
8	40	48.2	225	9	US-09-999-745-67
9	40	48.2	225	10	US-09-866-538-12
10	40	48.2	225	10	US-09-794-308-12
11	40	48.2	225	10	US-09-865-291-12
12	40	48.2	225	12	US-10-132-067-4
13	40	48.2	225	13	US-10-006-922-12
14	40	48.2	225	13	US-10-006-922-44
15	40	48.2	225	14	US-10-081-864-8

```

16 40 48.2 225 14 US-10-081-864-12 Sequence 12, Appli
17 40 48.2 225 14 US-10-121-258-1 Sequence 1, Appli
18 40 48.2 225 14 US-10-121-258-20 Sequence 20, Appli
19 40 48.2 225 14 US-10-315-920-2 Sequence 2, Appli
20 40 48.2 225 14 US-10-315-920-4 Sequence 4, Appli
21 40 48.2 225 15 US-10-370-570-56 Sequence 56, Appli
22 40 48.2 225 15 US-10-406-618-32 Sequence 32, Appli
23 40 48.2 225 16 US-10-433-640-13 Sequence 13, Appli
24 40 48.2 226 16 US-10-724-178-12 Sequence 12, Appli
25 40 48.2 230 13 US-10-006-922-18 Sequence 18, Appli
26 40 48.2 230 14 US-10-161-403-40 Sequence 40, Appli
27 40 48.2 236 16 US-10-314-936-2 Sequence 2, Appli
28 40 48.2 236 16 US-10-314-936-4 Sequence 4, Appli
29 40 48.2 240 14 US-10-152-296-2 Sequence 2, Appli
30 40 48.2 240 16 US-10-739-856-2 Sequence 112625,
31 40 48.2 326 16 US-10-437-963-112625 Sequence 1, Appli
32 40 48.2 487 15 US-10-343-977-1 Sequence 2, Appli
33 40 48.2 506 15 US-10-343-977-2 Sequence 52, Appli
34 40 48.2 545 14 US-10-214-932-52 Sequence 3, Appli
35 40 48.2 547 15 US-10-343-977-3 Sequence 76, Appli
36 40 48.2 548 14 US-10-214-932-76 Sequence 224455,
37 39 47.0 107 10 US-09-994-595-80 Sequence 67457, A
38 39 47.0 133 12 US-10-424-599-224455 Sequence 8, Appli
39 39 47.0 213 12 US-10-425-114-67457 Sequence 17, Appli
40 39 47.0 215 14 US-10-155-809-8 Sequence 41, Appli
41 39 47.0 222 15 US-10-406-618-17 Sequence 8, Appli
42 39 47.0 225 16 US-10-423-688A-41 Sequence 10, Appli
43 39 47.0 226 9 US-09-976-673-10 Sequence 26, Appli
44 39 47.0 226 9 US-09-976-673-10
45 39 47.0 226 9 US-09-976-673-26

```

ALIGNMENTS

```

RESULT 1
US-09-994-595-43
; Sequence 43, Application US/09994595
; Publication No. US20030039981A1
; GENERAL INFORMATION:
; APPLICANT: Bhattacharjee, J.
; APPLICANT: Suvarna, Kalavati
; APPLICANT: Bhattacharjee, Vasker
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; TITLE OF INVENTION: A BIOLOGICAL SAMPLE
; FILE REFERENCE: 96,247-A
; CURRENT APPLICATION NUMBER: US/09/994,595
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO 43
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide segment of ACVS_EMENI shown in Figure 4.
US-09-994-595-43

```

```

Query Match          51.8%; Score 43; DB 10; Length 107;
Best Local Similarity 50.0%; Pred No. 4.6;
Matches      8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      1 SVIAKQMTYKVYMSGT 16
       |:|:|:|:|:|
Db      1 SLTSKQLAYVTYTSQT 16

```

```

RESULT 2
US-10-424-599-172498
; Sequence 172498, Application US/10424599
; Publication No. US20040031072A1

```

```

; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 172498
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(47)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_126782C.1.pep
US-10-424-599-172498

```

```

Query Match          50.6%; Score 42; DB 12; Length 47;
Best Local Similarity 50.0%; Pred. No. 2.8;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 IAKQMTYKVMMSGT 16
   :|:|:|:|:|:|:|
Db 16 VARQPTIRIYMLGT 29

```

```

RESULT 3
US-10-315-920-6
; Sequence 6, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Pradkov, Arcady Fedorovich
; APPLICANT: Tetsikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077C1P
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-6

```

```

Query Match          49.4%; Score 41; DB 14; Length 225;
Best Local Similarity 52.9%; Pred. No. 25;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SVIAKQMTYKVMMSGTV 17
   :|:|:|:|:|:|:|
Db 6 NVITEFMRFKVRMEGTV 22

```

```

RESULT 4
US-10-442-148A-7
; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO

```

```

; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583.0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

```

```

Query Match          49.4%; Score 41; DB 15; Length 225;
Best Local Similarity 52.9%; Pred. No. 25;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SVIAKQMTYKVMMSGTV 17
   :|:|:|:|:|:|:|
Db 6 NVITEFMRFKVRMEGTV 22

```

```

RESULT 5
US-10-442-148A-8
; Sequence 8, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO
; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583.0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-8

```

```

Query Match          49.4%; Score 41; DB 15; Length 239;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SVIAKQMTYKVMMSGTV 17
   :|:|:|:|:|:|:|
Db 6 NVITEFMRFKVRMEGTV 22

```

```

RESULT 6
US-10-081-864-25
; Sequence 25, Application US/10081864
; Publication No. US2003002287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Pradkov, Arcady
; TITLE OF INVENTION: No. US2003002287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067

```

```

; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant fragment
US-10-081-864-25

```

```

Query Match      48.2%; Score 40; DB 14; Length 26;
Best Local Similarity 52.9%; Pred. No. 3.3;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 SVIAKQMTYKVMGTV 17
      :||:|:|||
Db      6 NVIKEFMRFKVRMEGTV 22

```

```

RESULT 7
US-10-006-922-46
; Sequence 46, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-46

```

```

Query Match      48.2%; Score 40; DB 13; Length 205;
Best Local Similarity 52.9%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 SVIAKQMTYKVMGTV 17
      :||:|:|||
Db      6 NVIKEFMRFKVRMEGTV 22

```

```

RESULT 8
US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent No. US20020157120A1

```

```

; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-999-745-67

```

```

Query Match      48.2%; Score 40; DB 9; Length 225;
Best Local Similarity 52.9%; Pred. No. 3.8;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 SVIAKQMTYKVMGTV 17
      :||:|:|||
Db      6 NVIKEFMRFKVRMEGTV 22

```

```

RESULT 9
US-09-866-538-12
; Sequence 12, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-866-538-12

```

```

Query Match      48.2%; Score 40; DB 10; Length 225;
Best Local Similarity 52.9%; Pred. No. 3.8;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 SVIAKQMTYKVMGTV 17
      :||:|:|||
Db      6 NVIKEFMRFKVRMEGTV 22

```

```

RESULT 10
US-09-794-308-12
; Sequence 12, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225

```



```
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hybrid construct
US-10-006-922-44
```

```
Query Match      48.2%; Score 40; DB 13; Length 225;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      1 SVIAKQMTYKYVMGTV 17
        :||:|:|:|:|:|:|
DB      6 NVIKFMRFKVRMEGIV 22
```

```
RESULT 15
US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication NO. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8
```

```
Query Match      48.2%; Score 40; DB 14; Length 225;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      1 SVIAKQMTYKYVMGTV 17
        :||:|:|:|:|:|:|
DB      6 NVIKFMRFKVRMEGIV 22
```

Search completed: August 12, 2004, 06:51:20
Job time : 35.7766 secs

This Page Blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	43	51.8	107	4	US-09-106-568E-43		Sequence 43, Appl
2	39	47.0	107	4	US-09-106-568E-80		Sequence 80, Appl
3	39	47.0	3665	2	US-08-222-617A-13		Sequence 13, Appl
4	39	47.0	3712	2	US-08-222-617A-4		Sequence 4, Appl
5	39	47.0	3712	2	US-08-222-617A-35		Sequence 25, Appl
6	38	45.8	195	4	US-09-489-039A-13455		Sequence 13455, A
7	38	45.8	209	4	US-09-134-001C-4478		Sequence 4478, Ap
8	37.5	45.2	329	4	US-09-540-236-3211		Sequence 3211, Ap
9	37	44.6	171	4	US-09-540-236-2461		Sequence 2461, Ap
10	37	44.6	351	4	US-09-198-452A-991		Sequence 991, App
11	36	43.4	390	4	US-09-488-039A-13547		Sequence 13547, A
12	36	43.4	431	4	US-09-107-532A-7056		Sequence 7056, Ap
13	35	42.2	28	1	US-08-446-692-3		Sequence 3, Appl
14	35	42.2	28	2	US-08-488-351A-3		Sequence 3, Appl
15	35	42.2	28	4	US-09-106-568E-21		Sequence 21, Appl
16	35	42.2	30	3	US-09-100-409A-42		Sequence 42, Appl
17	35	42.2	30	5	PCF-US95-13841-6		Sequence 6, Appl
18	35	42.2	38	1	US-08-446-692-11		Sequence 11, Appl
19	35	42.2	38	2	US-08-488-351A-11		Sequence 11, Appl
20	35	42.2	87	4	US-09-134-000C-5981		Sequence 5981, Ap
21	35	42.2	104	2	US-08-292-968-23		Sequence 23, Appl
22	35	42.2	104	2	US-08-467-974-23		Sequence 23, Appl
23	35	42.2	104	2	US-08-467-536-23		Sequence 23, Appl
24	35	42.2	104	3	US-08-467-976-23		Sequence 23, Appl
25	35	42.2	104	3	US-09-082-514-23		Sequence 23, Appl
26	35	42.2	107	4	US-09-106-568E-60		Sequence 60, Appl
27	35	42.2	174	4	US-09-252-991A-25897		Sequence 25897, A

```

; SOFTWARE: Microsoft Word 97
; SEQ ID NO 80
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypep
US-09-106-5683-80

```

Query Match	47.0%;	Score 39;	DB 4;	Length 107;
Best Local Similarity	53.8%;	Pred. No. 5.7;		
Matches	7:	Conservative	2:	Mismatches
			4:	Indels
			0:	Gaps
			0:	

QY	4	AKQMTYKVYMSGT	16
		: :	
pB	4	SKOLAXVTYTSGT	16

RESULT 3
US-08-222-617A-13
; Sequence 13, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Laempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase

Query Match 47.0%; Score 39; DB 2; Length 3665;
Best Local Similarity 53.8%; Pred. No. 3.4e+02;
Matches 7: Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AKQMTYKVYMSGT 16
:|:| | | | |
pb 414 SKOLAYVTTYTSGT 426

RESULT 4
US-08-222-617A-4
; Sequence 4, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

Query Match 47.0%; Score 39; DB 2; Length 3712;
Best Local Similarity 53.8%; Pred. NO. 3.4e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AKQMTYKVYMSGT 16
: || : | | | |
Db 414 SKOLAYVVTYTSGT 426

RESULT 5
US-08-222-617A-25
; Sequence 25, Application US/08222617A
; Patent No. 582879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.

Db 86 VLEKQMTSRGYMIG 99

RESULT 7

US-09-134-001C-4478

; Sequence 4478, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4478

; LENGTH: 209

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4478

Query Match 45.8%; Score 38; DB 4; Length 209;

Best Local Similarity 70.0%; Pred. No. 19;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 8 TYKYVMSGTV 17

|||||:|

Db 93 TYKYXNGVV 102

RESULT 8

US-09-540-236-3211

; Sequence 3211, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 3211

; LENGTH: 329

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-3211

Query Match 45.2%; Score 37.5; DB 4; Length 329;

Best Local Similarity 41.2%; Pred. No. 40;

Matches 7; Conservative 8; Mismatches 1; Indels 1; Gaps 1;

Qy 1 SVIAKQMTYKYVMSGTV 17

:::||||:::|

Db 154 TIVAKQITVQM-VNGNV 169

RESULT 9

US-09-540-236-2461

; Sequence 2461, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

```
; SEQ ID NO 2461
; LENGTH: 171
; TYPE: PRT
; ORGANISM: M.catarrhalis
; US-09-540-236-2461

Query Match      44.6%; Score 37; DB 4; Length 171;
Best Local Similarity 63.6%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      6 QMTYKVMVSGT 16
Db      28 QKTYKVIYIHT 38

RESULT 10
US-09-198-452A-991
; Sequence 991, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 991
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-991

Query Match      44.6%; Score 37; DB 4; Length 351;
Best Local Similarity 50.0%; Pred. No. 53;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      3 IAKQMTYKVMVSGT 16
Db      64 LARGMTYKALISNT 77

RESULT 11
US-09-489-039A-13547
; Sequence 13547, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13547
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13547

Query Match      43.4%; Score 36; DB 4; Length 390;
Best Local Similarity 46.7%; Pred. No. 93;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy      3 IAKQMTYKVMVSGT 17
Db      178 VAEFASQKVVYSGV 192

RESULT 12
US-09-107-532A-7056
; Sequence 7056, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7056:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...431
; SEQUENCE DESCRIPTION: SEQ ID NO: 7056:

Query Match      43.4%; Score 36; DB 4; Length 431;
Best Local Similarity 46.7%; Pred. No. 1e+02;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy      3 IAKQMTYKVMVSGT 17
Db      292 IIKVMTYSIMLGMV 306

RESULT 13
US-08-446-692-3
; Sequence 3, Application US/08446692
; Patent No. 5759551
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
```

```

; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,692
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)415-8745
; TELEFAX: (516)751-6849
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-446-692-3

```

```

Query Match 42.2%; Score 35; DB 1; Length 28;
Best Local Similarity 45.5%; Pred. No. 6.8;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 5 KOMTKYKVMMSG 15
   :::|:||||
Db 4 RLLYMIYMSG 14

```

```

RESULT 14
US-08-488-351A-3
; Sequence 3, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,351A
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,692
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/229,275
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/057,166
; FILING DATE: 27-APR-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)415-8745
; TELEFAX: (516)751-6849
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-488-351A-3

```

```

Query Match 42.2%; Score 35; DB 2; Length 28;
Best Local Similarity 45.5%; Pred. No. 6.8;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 5 KOMTKYKVMMSG 15
   :::|:||||
Db 4 RLLYMIYMSG 14

```

```

RESULT 15
US-09-106-568E-21
; Sequence 21, Application US/09106568E
; Patent No. 6455248
; GENERAL INFORMATION:
; APPLICANT: Bhattacherjee, J.
; APPLICANT: Suvarna, Kalavati
; APPLICANT: Bhattacherjee, Vasker
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; TITLE OF INVENTION: A BIOLOGICAL SAMPLE
; FILE REFERENCE: 96,247-A
; CURRENT APPLICATION NUMBER: US/09/106,568E
; CURRENT FILING DATE: 1998-06-29
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO 21
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide segment of ACVT_PENCH shown in Figure 4.
US-09-106-568E-21

```

```

Query Match 42.2%; Score 35; DB 4; Length 28;
Best Local Similarity 46.2%; Pred. No. 6.8;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 4 AKOMTKYKVMMSGT 16
   :::|:||||
Db 4 SSQLAYVTYTSGT 16

```

```

Search completed: August 12, 2004, 06:21:05
Job time : 2.77664 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 21.3012 Seconds
(without alignments)
1043.144 Million cell updates/sec

Title: US-09-890-463-3

Perfect score: 1268

Sequence: 1 SVIAKQMTYKYVMGTVNGH.....SIARKPLVACCFRVRKSRHK 231

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	184.5	14.6	238	1 JQ1514	green-fluorescent
2	106	8.4	583	2 S64909	probable membrane
3	90	7.1	752	2 S23818	hypothetical prote
4	87	6.9	787	2 E82323	organic solvent to
5	86.5	6.8	725	1 IUMSNG	neural cell adhesi
6	86	6.8	770	2 S00643	anthranilate synth
7	85	6.7	260	2 B70419	hypothetical prote
8	84	6.6	1162	2 S14939	E2 glycoprotein pr
9	83.5	6.6	390	2 D64903	probable arylsulfa
10	83.5	6.6	3461	2 S58870	reelin precursor -
11	82.5	6.5	876	2 A89944	alanyl-tRNA synth
12	82	6.5	248	2 S58096	hypothetical prote
13	82	6.5	1162	2 S14940	E2 glycoprotein pr
14	81.5	6.4	390	2 F90891	probable enzyme li
15	81.5	6.4	390	2 C85726	probable enzyme Z2
16	80	6.3	1822	2 S63985	hypothetical expor
17	79	6.2	334	2 D95982	collagen alpha 2 c
18	79	6.2	403	2 T51828	probable photosyst
19	79	6.2	687	1 B32382	ubiquinol-cytochro
20	78.5	6.2	363	2 S30149	cysteine proteinas
21	78.5	6.2	1254	2 S46636	hypothetical prote
22	77.5	6.1	266	1 CDFJ33	chlorophyll a/b-bi
23	77.5	6.1	269	1 G65102	probable transcrip
24	77.5	6.1	269	2 A91130	transcription repr
25	77.5	6.1	269	2 A85975	robable transcript
26	77.5	6.1	333	2 B64380	hypothetical prote
27	77.5	6.1	513	2 D59824	conserved hypothet
28	77.5	6.1	870	2 T47454	lipoxigenase AtLOX
29	77.5	6.1	896	2 JQ2391	lipoxigenase (EC 1

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 23-Mar-2001

C:Accession: J80692; JQ1514; PQ0335; S48693; S51330; S51331

R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.

Gene 111, 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: JQ1514; MUID:92175527; PMID:1347277

A:Accession: J80692

A:Molecule type: DNA

A:Residues: 1-107,'S',109-238 <PRA1>

A:Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99,'P',101-140,'L',142-218,'V',220-238 <PRA2>

A:Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>

A:Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'R',;

A:Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24,'Q',26-29,'R',31-83,'L',85-153,'G',155-156,'P',158-171,'K',173-208,'Q',;

A:Cross-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011

A:Experimental source: clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A',2-79,'R',81-9;

A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

protein kinase (EC
xylulokinase homol
WD-40 repeat prote
chlorophyll a/b-bi
hypothetical prote
pyruvate kinase (E
RNA-directed RNA p
El glycoprotein -
alpha-1-B-glycopro
hypothetical prote
hypothetical prote
hypothetical prote
genome polyprotein
structural polypro
chlorophyll a/b-bi
32.3K hypothetical

30 77.5 6.1 1702 2 T14050
31 77 6.1 487 2 E69895
32 76.5 6.0 1526 2 AC2239
33 76 6.0 266 1 CDFJ25
34 76 6.0 444 2 E65045
35 76 6.0 526 2 S26869
36 76 6.0 725 1 P21VBS
37 75.5 6.0 207 2 A4608
38 75.5 6.0 237 2 A42013
39 75.5 6.0 275 2 T24593
40 75.5 6.0 297 2 E71425
41 75.5 6.0 438 2 G71175
42 75.5 6.0 1241 2 S26373
43 75.5 6.0 1242 2 S72350
44 75 5.9 267 1 CDFJ2L
45 75 5.9 276 2 B86370

A:Gene: VC0446

A:Map position: 1

C:Superfamily: organic solvent tolerance protein

Query Match 6.9%; Score 87; DB 2; Length 787;
Best Local Similarity 19.2%; Pred. No. 12;
Matches 45; Conservative 33; Mismatches 92; Indels 64; Gaps 9;

QY 6 QMTYKVMSTGVN-----GHVFEVGGKGYEGEQTAVLTKGGLPFA--WDIL 56

Db 388 QLSYNYAPETMKYLLDLVSHVSREFTDARKP-----SATRVHPEGLKIFSNWGNW 443

QY 57 SPQCVQSGIPFTKYBEDIDPYVKQSPGGRYTWERIMNFDGAVCTVSNDSIQGNCFIYH 116

Db 444 TTEAR---VLGTYQODLDKTTD-----AKLEESVTRVPIERSV----- 480

QY 117 VKFSGNLFPNPGVMOKTQGWENPTELF-----RDGMLIGNFVAL----- 160

Db 481 ---AGIVLERDVTLLDDYDTQLEPKIQLYVPEKYQDNIGLYDSTLLQTDYYGLFRSKY 537

QY 161 ----KLEGG-----GHYLCFESTYKARKPKVMPGYHYVDKLDVTHNNDYTS 205

Db 538 SGVDRIESANQVSYGASTREFPDSNYKERLNIAGQIFLYLDSKLNPSNKNPDSITS 591

RESULT 5

IUMSNG

neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse

N:Alternate names: NCAM-120

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 31-Dec-2000

C:Accession: A29673; S00382; A44290

R:Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec

EMBO J. 6, 907-914, 1987

A:Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000

A:Reference number: A29673; MUID:87246524; PMID:3595563

A:Accession: A29673

A:Molecule type: mRNA

A:Residues: 1-725 <BAR>

A:Cross-references: EMBL:Y00051; NID:953342; PIDN:CAA68263.1; PID:953343

R:Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.

EMBO J. 7, 625-632, 1988

A:Title: Differential splicing and alternative polyadenylation generates distinct NCAM t

A:Reference number: S00382; MUID:86283628; PMID:3396534

A:Accession: S00382

A:Molecule type: DNA

A:Residues: 642-656, 'D' 658-725 <BA2>

A:Cross-references: EMBL:X07195

R:Rougou, G.; Marshak, D.R.

J. Biol. Chem. 261, 3396-3401, 1986

A:Title: Structural and immunological characterization of the amino-terminal domain of m

A:Reference number: A44290; MUID:86140120; PMID:3512556

A:Accession: A44290

A:Molecule type: protein

A:Residues: 20-36 <ROU>

C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol

C:Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IJMS

C:Genetics:

A:Gene: NCAM

A:Map position: 9

A:Introns: 701/1

C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu

C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane

F:1-19/Domain: signal sequence #status predicted <SIG>

F:132-191/Domain: immunoglobulin homology <IMM1>

F:152-156/Region: immunoglobulin homology <IMM2>

F:161-165/Region: heparin binding #status predicted

F:228-230/Domain: heparin binding #status predicted

F:263-272/Region: immunoglobulin homology <IMM3>

F:323-368/Domain: NCAM binding #status predicted

F:420-482/Domain: immunoglobulin homology <IMM4>

F:519-596/Domain: immunoglobulin homology <IMM5>

F: fibronectin type III repeat homology <FN3A>

F:625-685/Domain: fibronectin type III repeat homology <FN3B>

F:41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted

F:222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.8%; Score 86.5; DB 1; Length 725;
Best Local Similarity 21.7%; Pred. No. 12;
Matches 53; Conservative 41; Mismatches 89; Indels 61; Gaps 15;

QY 10 KYVMSTGVNHYFEVGGDKGK--PYE-----GEQTVR-LAVTKGGLPFAWDILSP- 58

Db 455 KII--NTPSASYLEVTDPSENDFGNYNCTAVNRIGQESLEFLIVQADTPSSPSIDRVPEY 512

QY 59 ----OCQY-----GSIPFTKVPEDIPDYVKQSPGGRYTWERIMNFDGAVCTVSNDSI 108

Db 513 SSTAQVQDPEPEATGVPIPKYKAWSKSLGESW--HFTWYD-----AKEANM 558

QY 109 QQNCFI-----YHVKFSGLNFPNGFVMO---KKTQG-----WEPNTERLPARDGMLI 153

Db 559 EGIVTIMGLKPETTYSDRLAALNGKLGHEIMQPSKSTQVPVPELSAPKLEGQWGEDGNSI 618

QY 154 GNNFMALKLEGG-----HYLCFESTYKARKP-VKMP--GYHYVDKLDVTHNNDYTSV 206

Db 619 KVNVL--IKODDGGSPIRHYLVKYLASEWKEFIRLPSGSHHMLKSLDWAAYEYVYVA 676

QY 207 EQRE 210

Db 677 ENQ 680

RESULT 6

S00643

anthranilate synthase multifunctional protein - Aspergillus niger

N:Alternate names: gene trpG-trpC-trpF protein; glutamine amidotransferase

N:Contains: anthranilate synthase (EC 4.1.3.27); indole-3-glycerol-phosphate synthase (EC

C:Species: Aspergillus niger

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 31-Mar-2000

C:Accession: S00643; A23979; B23979

R:Kos, T.; Kuijvenhoven, A.; Hensing, H.G.M.; Pouwels, P.H.; van den Hondel, C.A.M.J.J.

Curr. Genet. 13, 137-144, 1988

A:Title: Nucleotide sequence of the Aspergillus niger trpC gene: structural relationship

A:Reference number: S00643; MUID:88223483; PMID:2836085

A:Accession: S00643

A:Molecule type: DNA

A:Residues: 1-770 <KOS>

A:Cross-references: EMBL:X07071; NID:92420; PIDN:CAA30107.1; PID:92421

R:Kos, A.; Kuijvenhoven, J.; Wernars, K.; Bos, C.J.; van den Broek, H.W.J.; Pouwels, P.H.

Gene 39, 231-238, 1985

A:Reference number: A91539; MUID:86137391; PMID:2936650

A:Accession: A23979

A:Molecule type: DNA

A:Residues: 1-69 <KO2>

A:Cross-references: GB:M14404; NID:9166536; PIDN:AAA32709.1; PID:9166537

A:Accession: B23979

A:Molecule type: DNA

A:Residues: 392-433 <KO3>

A:Cross-references: GB:M14403; NID:9166538; PIDN:AAA32710.1; PID:9166539

C:Genetics:

A:Gene: trpC

C:Superfamily: trpG-trpC-trpF trifunctional enzyme; trpC homology; trpF homology; trpG ho

C:Keywords: carbon-carbon lyase; carboxy-lyase; intramolecular lyase; isomerase; oxo-acic

F:25-219/Domain: glutamine amidotransferase #status predicted <GAT>

F:26-216/Domain: trpG homology <TRG>

F:255-518/Domain: trpC homology <TRC>

F:255-514/Domain: indole-3-glycerol-phosphate synthase #status predicted <IGPS>

F:537-767/Domain: trpF homology <TRF>

F:545-770/Domain: N-[5'-phosphoribosyl]anthranilate isomerase #status predicted <PRAI>

F:104/Active site: Cys #status predicted

Query Match 6.8%; Score 86; DB 2; Length 770;

Best Local Similarity 24.3%; Pred. No. 15;

Matches 50; Conservative 23; Mismatches 61; Indels 72; Gaps 12;

QY 10 KYVMSTGVNHYFEVGGDKGKPYEGEQTAVLTKGGLPFAWDILSPQCYGSIPTFT 68

```

Db      117 KVDVTGELLHGKSPKXDKGK-AYEG-----LP-----GSLAVT 150
QY      69 KYPE-----DIPYVKQSPGRTYWRIMNFEDGACTVSNDSIOGNCFIYHVKFS-- 120
Db      151 RYHSLAGTHATIPDCLEVS-----SSVQLADD-----SNKDVIMG---VRHKKLAVE 194
QY      121 GLNFPNGPVQKKTQGWEPNTERLFARDGMLIGNFMALKL---EGGGHYLCFFKSTYK 177
Db      195 GVQPHF-----ESILTEYGRIMPRNFKLKTAGTWEGNGKHFGQSSSTTK 238
QY      178 ARKPKVKMPGYHYVDRKLDVTNNHKDY 203
Db      239 ATVPSNPP--PKTDKLSILERIYDH 262

RESULT 7
B70419
hypothetical protein aq_1369 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 04-Mar-2000
C:Accession: B70419
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70419
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-260 <AOF>
A:Cross-references: GB:AB000737; NID:92983782; PIDN:AAC07356.1; PID:92983798; GB:AE00065
A:Experimental source: strain VP5
C:Genetics:
A:Gene: aq_1369
C:Superfamily: Aquifex aeolicus hypothetical protein aq_1369

Query Match          6.7%; Score 85; DB 2; Length 260;
Best Local Similarity 22.3%; Pred. No. 4.6;
Matches 37; Conservative 28; Mismatches 71; Indels 30; Gaps 7;

QY      71 PEDIPYVKQSPGRTYWRIMNFEDGACTVSNDSIOGNCFIYHVKFSGLNPPP 126
Db      46 PENVREFLKENYPERKLIENWEELQGBFD-----VQKLGNEVLIVIRIPEKEPEK 97

QY      127 NGPVNKKTKQ-----GWEP--PNTERLFARDGMLIGNFM-ALKLEGGGHYLCE 171
Db      98 ELGIFQSVEEAMGAFSLTALEHGWEEVPKNYVIYHADFVEGNGKLIATAIKTEEGISTYDQ 157

QY      172 FKSTYKARKPKVMPGYYVDRKLDVTNNKQ--YTSVEQREISIAK 216
Db      158 LKLEEMKKNVRYP--RVVVSSDVLTYIKDIYDPVQSKAYVIARE 201

RESULT 8
S14939
E2 glycoprotein precursor - avian infectious bronchitis virus (strain M42-H)
N:Alternate names: peplomer glycoprotein
C:Species: avian infectious bronchitis virus, IBV
A:Variety: strain M42-H
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
C:Accession: S14939
R:Niesters, H.G.M.; Lenstra, J.A.; Spaan, W.J.M.; Zijderveld, A.J.; Bleumink-Pluym, N.M.
Virus Res. 5, 253-263, 1986
A:Title: The peplomer protein sequence of the M41 strain of coronavirus IBV and its comp
A:Reference number: S07421; MUID:87021475; PMID:2429473
A:Accession: S14939
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: genomic RNA
A:Residues: 1-1162 <NIE>
C:Superfamily: coronavirus E2 glycoprotein

Query Match          6.6%; Score 84; DB 2; Length 1162;

```

```

Best Local Similarity 21.1%; Pred. No. 37;
Matches 61; Conservative 32; Mismatches 62; Indels 134; Gaps 17;

QY      3 IAKQMTYK-----VYMSG-----TVNGHYFEVEGDGKGKPYEGEQTVR 40
Db      149 VAKYPTFRSQCVNLLTSVYNGDLVYTSNETDVTDSAGYVFKA----- 192
QY      41 LAVTKGGPLPF-----AWDIL-----SPQ-----COYGSIPFTK--YPEDI 74
Db      193 -----GGPIYKVMREVKALAYFVNGTAQDVILCDGSPRGLLACQYNTGNFSDGFYPTN 247
QY      75 PDYVKQSPGRTYWRIMNFEDGAV---CTVSNDSIOGNCFIYHVKFSGLNFPNG--- 128
Db      248 SSLVKQKF-----IVYRENSVNTTCTLHN-----FIPHNE-TGANPNPSPGVQN 289
QY      129 -PYMQKKT-QGWEPNTERLFARDGMLIGNFMALKLEGGGHYLCEFKSTYKARKPKVMPG 186
Db      290 IQYTQTKTAGSYNYFNFSFLSSFYKESNFM-----VGSYHPSCKF----- 331
QY      187 YHYVDRKLDVTNNHKDYTSVEQREISIAKPKL-----VACCF 223
Db      332 -----RLETINLGLWFNSL---SVSIAYGLOGGCKQSVFKGRATCCY 371

RESULT 9
D64903
probable arylsulfatase activating enzyme (EC 1.-.-.-) b1497 [similarity] - Escherichia
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64903
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64903
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-390 <BLAT>
A:Cross-references: GB:AE000247; GB:U00096; NID:g1787773; PIDN:AAC74570.1; PID:g1787774;
A:Experimental source: strain K-12, substrain MGL1655
C:Genetics:
A:Start codon: TTG
C:Superfamily: arylsulfatase activating enzyme atsb
C:Keywords: oxidoreductase

Query Match          6.6%; Score 83.5; DB 2; Length 390;
Best Local Similarity 19.6%; Pred. No. 10;
Matches 55; Conservative 46; Mismatches 104; Indels 75; Gaps 12;

QY      10 KYVMSTGVNGHYFEVEGDGKGK-----YEGEQTVRLAV-TKGGPLPF 51
Db      49 KOYIAASGNQVYFTWQG---GEPTLAGLDFFRKVTHYQORYAGOKRIFNALQTNGLNN 105
QY      52 AWDILSPQCYGSIPTFKYKPEDIPYVKQSPGRTYWRIMNFEDGACTVSNDSIOGNCFIYHVKFSGLNFPNG--- 103
Db      106 ENCAFLKEHEFTLVGISIDGPQELHRYRNRSGNFTFAKVIATAIERLSYQVEFNLTVT 165
QY      104 NDSSIQGNCFIVH-----VKFSG-----LNFPN--GP 129
Db      166 NNVNHYEVLVHFKLSGSKHMQFIELLETCTPNIDFSGHSENTRIIDFSVPYATYK 225
QY      130 VMQKKTQGWEPN-TERLFARDGMLIGNFMALKLEGGGHYLCEFKSTYKARKPKVMPG-- 186
Db      226 FMSTIFMQWKNVDGVEIFITQEPESVSRFL-----GNHTSCIFQESCKDLNIVESNGDI 280
QY      187 ----YHYVDRKLDVTNNK--DYTSVEQREISIAKPKLVACC 222
Db      281 YECDFHVPQYKIGKINKSELKTMNSVOLTAOKKRIPAKC 320

RESULT 10
S58870

```


Query Match 6.5%; Score 82.5; DB 2; Length 876;
Best Local Similarity 21.9%; Pred. No. 35;

A; Variety: strain M42-S

RESULT 13
S14940
E2 glycoprotein precursor - avian infectious bronchitis virus (strain M42-S)
N:Alternate names: peplomer glycoprotein
C:Species: avian infectious bronchitis virus, IBV
A:Variety: strain M42-S

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 14.2008 Seconds
(without alignments)
847.008 Million cell updates/sec

Title: US-09-890-463-3

Perfect score: 1288

Sequence: 1 SVIAKQMTYKVMSTGVNH.....SIARKPLVACCFRVRKSRHK 231

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	187.5	14.8	238	1	GFP_AEQVI
2	87	6.9	787	1	OSTA_VIBCH
3	86.5	6.8	725	1	NCA2_MOUSE
4	86	6.8	770	1	TRPG_ASPNG
5	85	6.7	260	1	YD69_AQUAE
6	84	6.6	1162	1	VGL2_IBVB
7	83.5	6.6	385	1	YDEM_ECOLI
8	83.5	6.6	3461	1	RELN_MOUSE
9	82.5	6.5	876	1	SYA_STAAM
10	82.5	6.5	876	1	SYA_STAAM
11	82	6.5	248	1	YAI4_SCHPO
12	81.5	6.4	3462	1	RELN_RAT
13	81	6.4	1142	1	ENAM_HUMAN
14	79.5	6.3	994	1	ATAL_CHICK
15	79	6.2	403	1	H136_ARATH
16	79	6.2	687	1	CYBC_BRAJA
17	78.5	6.2	574	1	IRL2_MOUSE
18	78.5	6.2	1254	1	UBPC_YEAST
19	77.5	6.1	266	1	CB21_PETSP
20	77.5	6.1	269	1	AGAR_ECOLI
21	77.5	6.1	333	1	Y642_METJA
22	77.5	6.1	366	1	MLTA_BUCAP
23	77.5	6.1	513	1	YHCX_BACSU
24	77.5	6.1	896	1	LOXC_ARATH
25	77	6.1	726	1	RRP2_INBP9
26	76.5	6.0	1526	1	YI46_ANASP
27	76	6.0	266	1	CS24_PETSP
28	76	6.0	444	1	YGAF_ECOLI
29	76	6.0	725	1	RRP2_INBSI
30	75.5	6.0	1960	1	TF20_HUMAN
31	75	5.9	267	1	CB22_PETSP
32	75	5.9	277	1	YF99_METJA
33	75	5.9	342	1	XYNA_CALSA

34 75 5.9 392 1 ANIA_NEIGO Q02219 neisseria g
35 75 5.9 809 1 OSTA_XANCP Q8pce0 xanthomonas
36 75 5.9 1983 1 TF20_MOUSE Q9epq8 mus musculus
37 74.5 5.9 587 1 COAT_PAVL3 P36310 parvovirus
38 74.5 5.9 2390 1 SPCP_HUMAN O15020 homo sapien
39 74 5.8 265 1 CB2B_LYCES P07370 lycopersico
40 74 5.8 267 1 CB25_PETSP P04783 petunia sp.
41 74 5.8 398 1 Y4QJ_RHISN P55631 rhizobium s
42 74 5.8 443 1 GAT3_MOUSE P23772 mus musculus
43 74 5.8 813 1 OSTA_XANAC Q8p222 xanthomonas
44 74 5.8 876 1 SYA_STAEP Q8csa7 staphylococ
45 74 5.8 1004 1 MV10_MOUSE P23249 mus musculus

ALIGNMENTS

RESULT 1
GFP_AEQVI
ID GFP_AEQVI STANDARD; PRT; 238 AA.
AC P4212; Q17104; Q27903;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein."
RL Gene 111:229-233 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein."
RL FEBS Lett. 341:277-280 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage."
RL Plant Mol. Biol. 33:989-999 (1997).
RN [4]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein."
RL Biochemistry 32:1212-1218 (1993).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RP MEDLINE=96355665; PubMed=8703075;
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein."
RL Science 273:1392-1395 (1996).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RP MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein."
RL Nat. Biotechnol. 14:1246-1251 (1996).

[7]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.
 RX MEDLINE=98455509; PubMed=9782051;
 RA Wachter R.M., Elsiger M.A., Kallio K., Hanson G.T., Remington S.J.;
 RT "Structural basis of spectral shifts in the yellow-emission variants
 of green fluorescent protein."; Structure 6:1267-1277(1998).
 RL [8]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=99238303; PubMed=10220315;
 RA Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 RT "Structural and spectral response of green fluorescent protein
 variants to changes in pH."; Biochemistry 38:5236-5301(1999).
 RL [9]
 CC -1- FUNCTION: Energy-transfer acceptor. Its role is to transduce the
 blue chemiluminescence of the protein aequorin into green
 fluorescent light by energy transfer. Fluoresces in vivo upon
 receiving energy from the Ca(2+)-activated photoprotein aequorin.
 Absorbs light maximally at 395 nm and exhibits a smaller
 absorbance peak at 470 nm. The fluorescence emission spectrum
 peaks at 509 nm with a shoulder at 540 nm.
 CC -1- SUBUNIT: Monomer.
 CC -1- TISSUE SPECIFICITY: Photocytes.
 CC -1- PTM: Contains a covalently attached chromophore, which is composed
 of modified amino acid residues. The chromophore is formed upon
 cyclization of the residues Ser-dehydroTy-Gly.
 CC -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
 chimeric proteins of GFP linked to other proteins where it
 functions as a fluorescent protein tag. GFP tolerates N- and C-
 terminal fusion to a broad variety of proteins. It has been
 expressed in bacteria, yeast, slime mold, plants, Drosophila,
 zebrafish, and in mammalian cells. As a noninvasive fluorescent
 marker in living cells, it allows for a wide range of applications
 where it may function as a cell lineage tracer, reporter of gene
 expression, or as a measure of protein-protein interactions.
 CC -1- DATABASE: NAME-Protein Spotlight;
 NOTE=Issue 11 of June 2001;
 WWW="http://www.expasy.org/spotlight/articles/sptl011.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M62654; AAA27722.1; -;
 CC EMBL; M62653; AAA27721.1; -;
 CC EMBL; L29345; AAA58246.1; -;
 CC EMBL; X96418; CAA65278.1; -;
 CC PIR; J80692; JQ1514.
 CC PDB; 1B9C; 17-NOV-00.
 CC PDB; 1BFP; 07-JUL-97.
 CC PDB; 1C4F; 14-JUN-00.
 CC PDB; 1EMA; 08-NOV-96.
 CC PDB; 1EMB; 16-JUN-97.
 CC PDB; 1EMC; 20-AUG-97.
 CC PDB; 1EME; 20-AUG-97.
 CC PDB; 1EMF; 20-AUG-97.
 CC PDB; 1EMG; 12-MAY-99.
 CC PDB; 1EMK; 20-AUG-97.
 CC PDB; 1EML; 20-AUG-97.
 CC PDB; 1EMM; 20-AUG-97.
 CC PDB; 1F09; 17-NOV-00.
 CC PDB; 1F0B; 11-JAN-97.
 CC PDB; 1GFL; 11-JAN-97.
 CC PDB; 1HCJ; 15-JAN-02.
 CC PDB; 1HUY; 04-JUL-01.
 CC PDB; 1JBY; 07-JAN-03.
 CC PDB; 1JEB; 07-JAN-03.
 CC PDB; 1KP5; 28-AUG-02.
 CC PDB; 1KYP; 10-APR-02.

DR PDB; 1KYR; 10-APR-02.
 DR PDB; 1KYS; 10-APR-02.
 DR PDB; 1YFP; 28-OCT-98.
 DR PDB; 2EMD; 20-AUG-97.
 DR PDB; 2EMN; 20-AUG-97.
 DR PDB; 2EMO; 20-AUG-97.
 DR InterPro; IPR000917; GFP_like.
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP; 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 KW Luminescence; 3D-structure.
 FT CROSSLINK 65 67 5-imidazolinone (Ser-Gly).
 FT MOD RES 66 66 2,3-DIDEHYDROTYROSINE.
 FT VARIANT 100 100 F -> Y.
 FT VARIANT 108 108 T -> S.
 FT VARIANT 141 141 L -> M.
 FT VARIANT 219 219 V -> I.
 FT CONFLICT 2 2 S -> G (IN REF. 3).
 FT CONFLICT 25 25 H -> Q (IN REF. 2).
 FT CONFLICT 80 80 Q -> R (IN REF. 3).
 FT CONFLICT 157 157 Q -> P (IN REF. 2).
 FT CONFLICT 172 172 E -> K (IN REF. 2).
 FT HELIX 4 8
 FT STRAND 12 22
 FT TURN 23 24
 FT STRAND 25 36
 FT TURN 37 40
 FT STRAND 41 48
 FT TURN 49 50
 FT HELIX 57 60
 FT TURN 61 63
 FT HELIX 69 71
 FT STRAND 73 73
 FT HELIX 76 81
 FT HELIX 83 86
 FT TURN 87 90
 FT STRAND 92 100
 FT TURN 101 102
 FT STRAND 105 115
 FT TURN 116 117
 FT STRAND 118 128
 FT TURN 132 133
 FT TURN 135 139
 FT STRAND 141 141
 FT STRAND 148 155
 FT TURN 156 159
 FT STRAND 160 171
 FT TURN 172 173
 FT STRAND 176 187
 FT STRAND 199 208
 FT TURN 211 212
 FT STRAND 217 227
 SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21F8FB6E05 CRC64;
 Query Match 14.8%; Score 187.5; DB 1; Length 238;
 Best Local Similarity 25.2%; Pred. No. 9.5e-10;
 Matches 52; Conservative 46; Mismatches 91; Indels 17; Gaps 8;
 QY 11 VMSGTVNGHYFEVGGDGKPKYEGEQTVLAVTKGGLPFAWDILSPQCOVGSIPFTKY 70
 Db 16 VELDGDVNGHKFSVSGEGEDATYTKLTKICTT-GKLPPVPTLVTTFTSYGVQCFERY 74
 QY 71 PEDIP--DYVQKSPFGRYTWERIMNFEDGAVCTVSNDSSIQNCFIYHKFSGLNFPNG 128
 Db 75 PDHMKQHDFFKSAPEGVQERTIFFDDGNGYKTRAEVKFEGDTLVNRILKIDFKEDG 134
 QY 129 PVMQKKTGQWEPNTERLF-----ARDGMLIGNNF-MALKLEGGHYLCEF--KSYTKARK 180
 Db 135 NILGHKLE-YNYSNHNVTIMADKQNGIKV--NFKIRHNIEDGVSVOLADHYQOQNTPIGDG 191
 QY 181 PVKMGYHYVDKLDVLT--NHNKDY 203

```

Db 192 PVLDPNHYLSQTSALSCKDPNEKRHD 217
RESULT 2
OSTA_VIBCH STANDARD; PRT; 787 AA.
ID Q9KUR9;
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Organic solvent tolerance protein precursor.
GN IMP OR OSTA OR VCU446.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: Determines N-hexane tolerance. Involved in outer
CC membrane permeability. Essential for envelope biogenesis. Could be
CC part of a targeting/usher system for outer membrane components (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- SIMILARITY: Belongs to the imp/osta family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: AE004131; AAF93619.1; --
CC PIR: E82323; E82323.
CC TIGR: VC0446; --.
CC HAVAP: MF_01411; -.
CC InterPro: IPR005653; Osta.
CC InterPro: IPR007543; Osta_C.
CC Pfam: PF03968; Osta_A; 1.
CC Pfam: PF04453; Osta_C; 1.
CC Outer membrane; Signal; Complete proteome.
KW SIGNAL
FT CHAIN 29 787 ORGANIC SOLVENT TOLERANCE PROTEIN.
SQ SEQUENCE 787 AA; 89017 MW; 036718F1896E0F7D CRC64;
Query Match 6.9%; Score 87; DB 1; Length 787;
Best Local Similarity 19.2%; Pred. No. 4.4;
Matches 45; Conservative 33; Mismatches 92; Indels 64; Gaps 9;
Qy 6 QMTYKVMYMGTVN-----GHYFEVGDGKPKVEGEQTVRVLAVTGGPLPEA--WDIL 56
Db 388 QLSNYVAPETMKYLDLVLVSHVSREFTDAGKP-----SATRVHIEPLKIPFNTWGNW 443
Qy 57 SPQCYGSIPTKYPIDIPYVKGSFFGRYTWIRIMNFEDGAVCTVSNDDSIQNCPIYH 116
Db 444 TTEAR---VLGTYYQQLDKTTD-----AKLESVTRVIEIRSV----- 480
Qy 117 VKFGLNFPNPGFVWQKKTQGWENTERLFA-----RDQMLIGNFMAL----- 160
Db 481 ---AGVLERDVTLLDDYTQTLEPKIQVLYVPEKYQDNIGLYDSTLLQTDYYGLFRSKY 537
QY 161 ----KLEGG-----GHYLCPEKSYTKARKPVKMPGHHVDRKLDVTNNKQVTS 205
Db 538 SGVDRIESANQVSYGASTRFFDSNYKERLNTAFQGIIFYLDSKLNFSKNKPDSTS 591
RESULT 3
ID NCA2_MOUSE STANDARD; PRT; 725 AA.
AC P13594; Q61950;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE (NCAM-120).
GN NCAM1 OR NCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=87246524; PubMed=3595563;
RA Barthels D., Santoni M.-J., Wille M., Ruppert C., Caix J.-C.,
RA Hirsch M.-R., Pontecilla-Camps J.-C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT a Mr 79,000 polypeptide without a membrane-spanning region.";
RL EMBO J. 6:907-914 (1987).
RN [2]
SEQUENCE OF 20-700 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=89251563; PubMed=2721486;
RA Santoni M.-J., Barthels D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
RT generates at least eight types of NCAM cDNA in mouse brain.";
RL EMBO J. 8:385-392 (1989).
RN [3]
SEQUENCE OF 642-725 FROM N.A.
RX MEDLINE=88283628; PubMed=3396534;
RA Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT "Differential splicing and alternative polyadenylation generates
RT distinct NCAM transcripts and proteins in the mouse.";
RL EMBO J. 7:625-632 (1988).
RN [4]
SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401 (1986).
CC -!- FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=N-CAM 120;
CC IsoId=P13594-1; Sequence=Displayed;
CC Name=N-CAM 180;
CC IsoId=P13595-1; Sequence=External;
CC Name=N-CAM 140;
CC IsoId=P13595-2; Sequence=External;
CC -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
CC EMBL; Y00051; CAA68263.1; -
DR EMBL; X15049; CAA33148.1; ALT_SEQ.
DR EMBL; X07195; CAA30173.1; -.
DR PIR; A29673; IQMSNG.
DR PDB; 2NCW; 12-MAR-97.
DR PDB; 3NCW; 23-JUL-99.
DR MGD; MGI:97281; Ncam1.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS00835; IG_LIKE; 5.
DR Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KW Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor;
FT SIGNAL 1 19
FT CHAIN 20 725
FT NEURAL CELL ADHESION MOLECULE 1, 120 kDa
FT ISOFORM.
FT DOMAIN 20 111
FT IG-LIKE C2-TYPE 1.
FT DOMAIN 116 205
FT IG-LIKE C2-TYPE 2.
FT DOMAIN 212 302
FT IG-LIKE C2-TYPE 3.
FT DOMAIN 309 402
FT IG-LIKE C2-TYPE 4.
FT DOMAIN 407 492
FT IG-LIKE C2-TYPE 5.
FT DOMAIN 519 596
FT FIBRONECTIN TYPE-III 1.
FT DOMAIN 625 692
FT FIBRONECTIN TYPE-III 2.
FT DOMAIN 152 156
FT HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 161 165
FT HEPARIN-BINDING (POTENTIAL).
FT DISULFID 41 96
FT PROBABLE.
FT DISULFID 139 189
FT PROBABLE.
FT DISULFID 235 288
FT PROBABLE.
FT DISULFID 330 386
FT PROBABLE.
FT DISULFID 427 480
FT PROBABLE.
FT CARBOHYD 222 222
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 348 348
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 424 424
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 450 450
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 479 479
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 261 268
FT ERSRVS -> DEKHIFSD (IN REF. 2).
FT CONFLICT 273 273
FT Q -> L (IN REF. 2).
FT CONFLICT 354 355
FT Q -> K (IN REF. 2).
FT CONFLICT 549 549
FT T -> R (IN REF. 2).
FT CONFLICT 572 572
FT D -> V (IN REF. 2).
FT CONFLICT 575 575
FT MOPSSES -> SAATEF (IN REF. 2).
FT CONFLICT 589 594
FT PEL -> REP (IN REF. 2).
FT CONFLICT 600 602
FT H -> D (IN REF. 2 AND 3).
FT CONFLICT 657 657
FT SEQUENCE 725 AA; 80296 MW; CAAEB8B4461C6B2F CRC64;
QY Match 6.8%; Score 86.5; DB 1; Length 725;
Best Local Similarity 21.7%; Pred. No. 4.4;
Matches 53; Conservative 41; Mismatches 89; Indels 61; Gaps 15;
QY 10 KYMSGTVNGHYFEVGGDKK--PYE-----GEQTVR-LAVTKGGLPFAMWILSP- 58
DB 455 KYY--NTFSASYLEVPDSENDGNYNCTAVNRIGQSELEFILVQADTFSPFSIDRVPEY 512
QY 59 ----QCQV-----GSIPFTKVPDIPYVKQSPFRGTWRIMNFEDGAVCTVSNDSI 108
DB 513 SSTAQVQDEPEATQGVILKYKAWSKSLGESW--HTTWYD-----AKEANM 558
QY 109 QGNCFI-----YHVFSGLNFPPGVNQ---KKTQG-----WEPNTERLFARDGMLI 153
DB 559 EGIIVTMGLKPTTYSDRLLAALNGKGLGEIMQPSSEKTPQVPFELSAKLEGGMGEDGNSI 618
QY 154 GNNFVALKLEGG-----HYLCEFKSTYKARP-VKMP--GYHYVDKLDVTNHNKDYTSV 206
DB 619 KYNL--IKODDGGSPIRHYLVKRYALASEWKEPELPSGSHHVMUKSLDWNAYEYVVA 676
```

```
QY 207 EQRE 210
DB 677 ENQ 680
RESULT 4
TRPG ASPNG STANDARD; PRT; 770 AA.
ID TRPG ASPNG STANDARD; PRT; 770 AA.
AC P05328;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Anthranilate synthase component II (EC 4.1.3.27) [Includes: Glutamine
DE amidotransferase; Indole-3-glycerol phosphate synthase (EC 4.1.1.48)
DE (IGPS); N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24)
DE (PRAI)].
DE TRPC.
GN Aspergillus niger.
OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichomaceae; mitosporic Trichomaceae; Aspergillus.
OC NCBI_TaxID=5061;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=401;
RC MEDLINE=88223483; PubMed=2836085;
RA Kos T., Kuijvenhoven A., Hensing H.G.M., Pouwels P.H.,
RA van den Hondel C.A.M.J.J.;
RT "Nucleotide sequence of the Aspergillus niger trpC gene: structural
RT relationship with analogous genes of other organisms.";
RL Curr. Genet. 13:137-144(1988).
RN [2]
RP SEQUENCE OF 1-69 AND 392-433 FROM N.A.
RX MEDLINE=8613791; PubMed=2936650;
RA Kos A., Kuijvenhoven J., Wernars K., Bos C.J., van den Broek H.W.J.,
RA Pouwels P.H., van den Hondel C.A.M.J.J.;
RT "Isolation and characterization of the Aspergillus niger trpC gene.";
RL Gene 39:231-238(1985).
CC -!- FUNCTION: TRIFUNCTIONAL ENZYME BEARING THE GLN AMIDOTRANSFERASE
CC (GATASE) DOMAIN OF ANTHRANILATE SYNTHASE, INDOLE-GLYCEROLPHOSPHATE
CC SYNTHASE, AND PHOSPHORIBOSYLANTHRANILATE ISOMERASE ACTIVITIES.
CC -!- CATALYTIC ACTIVITY: N-(5-phospho-beta-D-riboseyl)-anthranilate = 1-
CC (2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate.
CC -!- CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-
CC phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
CC -!- CATALYTIC ACTIVITY: Chorismate + L-glutamine = anthranilate +
CC pyruvate + L-glutamate.
CC -!- PATHWAY: Tryptophan biosynthesis; first step.
CC -!- PATHWAY: Tryptophan biosynthesis; third step.
CC -!- PATHWAY: Tryptophan biosynthesis; fourth step.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07071; CAA30107.1; -.
CC EMBL; M14403; AAA32710.1; -.
CC EMBL; M14404; AAA32709.1; -.
CC PIR; S00643; S00643.
CC HSP; Q06129; IQDL.
CC InterPro; IPR006220; Anth_synthII.
CC InterPro; IPR001317; CP_synthGatase.
CC InterPro; IPR003009; FMN_enzyme.
CC InterPro; IPR000991; Gatase_1.
CC InterPro; IPR001468; IGPS_1.
CC InterPro; IPR001240; PRAI.
CC InterPro; IPR006221; TrpG_papA.
CC Pfam; PF00117; Gatase; 1.
```

DR Pfam; PF00218; IGPS; 1.
DR Pfam; PF00697; PRAI; 1.
DR PRINTS; PR00097; ANTSENTHASRII.
DR PRINTS; PR00099; CPSCATASE.
DR PRINTS; PR00096; GATASE.
DR PRODOM; PD001511; IGPS; 1.
DR TIGRFAMs; TIGR00566; trpG_papA; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
DR PROSITE; PS00614; IGPS; 1.
KW Triptophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
KW Decarboxylase; Transferase; Glutamine amidotransferase.
FT DOMAIN 25 219
FT DOMAIN 255 519
FT DOMAIN 535 770
FT ACT_SITE 104 104
FT ACT_SITE 199 199
FT ACT_SITE 201 201
FT ACT_SITE 201 201
SQ SEQUENCE 770 AA; 82309 MW; 36D8D5B23097012 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 770;
Best Local Similarity 24.3%; Pred. No. 5.3;
Matches 50; Conservative 23; Mismatches 61; Indels 72; Gaps 12;

QY 10 KYVMGTV-NGHYFEVGEGRGKPYEGEQTVRLAVTKGGLPFAWDILSPQCQYGSIPFT 68
DB 117 KVDVTGEILHCKTSPLKDHGK-AYEG-----LP-----GSLAVT 150
QY 69 KYPE-----DIPYVKOSFFGRYTWIRIMPFEDGAVCTVNDSSIQNCFTYHVKFS-- 120
DB 151 RYHSLAGTHATIPDCLEVS-----SSVQLADD-----SNKQVIMG---VRHKKLAVE 194
QY 121 GLNPPPGVGVQKKTQGWEPNTERLFARDGMLGNFNMAKL---EGGHVLCERKSTYK 177
DB 195 GVQFHP-----ESLTYGRIMFRNFKLGTAGTWEGNGKHFGQSSTK 238
QY 178 ARKPVKMPGVHYVDRKLDVTNNKDY 203
DB 239 ATVPSNPP--PKTKKLSILRIYDH 262

RESULT 5
ID YD69 AQUAE STANDARD; PRT; 260 AA.
AC 067381;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein AQ_1369.
GN AQ_1369.
OS Aquifex aeolicus.
OC Bacteria; Aquificae;
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RA MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358 (1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000737; AAC07356.1; -.

DR PIR; B70419; B70419.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 260 AA; 30206 MW; B5D3FB6F37C89BB3 CRC64;

Query Match 6.7%; Score 85; DB 1; Length 260;
Best Local Similarity 22.3%; Pred. No. 1.8;
Matches 37; Conservative 28; Mismatches 71; Indels 30; Gaps 7;

QY 71 PEDPDYVKQSPGRY-----TWIRIMPFEDGAVCTVNDSSIQNCFTYHVKFSGLNFPF 126
DB 46 FENREFLNKPYEKYKLIENWEEIQGEFD-----VQKLGNGEYLIVIRIPEKEFEK 97
QY 127 NGPYVQKKTQ-----GWE-----PNTERLFARDGMLGNFNMAKLGNFNMAKLGNFNMAKL 171
DB 98 ELGIFQSVVEEAMGAPLSTALEHGWEEVKNYVIVHADVFEGGNKLIATKTEEGISTYDQ 157
QY 172 FKSTYKARKPVKMPGVHYVDRKLDVTNNKDY-----YTSVEQREISIAK 216
DB 158 LKLEEMKMKVRYR--RVVVSSDVLTYKDIYDPVQSKAYVIARE 201

RESULT 6
VGL2 IBVB
ID VGL2 IBVB STANDARD; PRT; 1162 AA.
AC PL1223; P05134;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE [Contains: Spike protein S1; Spike protein S2].
GN S.
OS Avian infectious bronchitis virus (strain Beaudette) (IBV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=111122;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85159540; PubMed=2984314;
RA Binns M.M., Boursnell M.E.G., Tomley F.M., Brown T.D.K.;
RA Brown T.D.K.;
RT "Cloning and sequencing of the gene encoding the spike protein of the
RT coronavirus IBV";
RL J. Gen. Virol. 66:719-726 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87085499; PubMed=3025348;
RA Binns M.M., Boursnell M.E.G., Tomley F.M., Brown T.D.K.;
RT "Comparison of the spike precursor sequences of coronavirus IBV
RT strains M41 and 6/82 with that of IBV Beaudette";
RL J. Gen. Virol. 67:2825-2831 (1986).
CC -!- FUNCTION: THE PEPLMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95169; AAA70235.1; -.
DR EMBL; X02342; CAA26201.1; -.
DR PIR; S14939; S14939.
DR InterPro; IPR002551; Corona_S1.
DR InterPro; IPR002552; Corona_S2.
DR Pfam; PF01600; Corona_S1; 1.
DR Pfam; PF01601; Corona_S2; 1.
KW Glycoprotein; Envelope protein; Transmembrane; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1162 E2 GLYCOPROTEIN.
FT CHAIN 19 537 SPIKE PROTEIN S1.
FT CHAIN 538 1162 SPIKE PROTEIN S2.

```
FT DOMAIN 1120 1137 CYS-RICH.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 247 247 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 306 306 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 979 979 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1014 1014 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1038 1038 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1051 1051 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1162 1162 AA; 128046 MW; 0BAADS9113C8EBD5 CRC64;
SQ SEQUENCE 1162 AA; 128046 MW; 0BAADS9113C8EBD5 CRC64;

Query Match 6.6%; Score 84; DB 1; Length 1162;
Best Local Similarity 21.1%; Pred. No. 13;
Matches 61; Conservative 32; Mismatches 62; Indels 134; Gaps 17;

Qy 3 IAKQMTYK-----VYMSG-----TNGHYFEVGGKGPYEGQTVR 40
Db 149 VAKYPTFRSFQVNNLTSLVNLGDLVYTSNETIDVTSGVYFKA-----192

Qy 41 LAVTKGGPLPF-----AWDL-----COYGSIPFTK--YPEDI 74
Db 193 -----GGPITTKVMREVKALAYFVNGTAQVILCDGSPRGLLACQYNTGNSDGFYFTN 247

Qy 75 PDYVKQSPGGRYTWERIMNFDGAV---CTVSNDSIQQNCFIYHVKFSGLNFPNG---128
Db 248 SSLVKQKF-----IVYRENSVNTTCTLHN-----FIEHNE-TGANPNPESGVQN 289

Qy 129 -PVMQKKT-QGWENPTEFLFARDGMLIGNFMALKLEGGHYLCFEFKSYTKARKPVKMPG 186
Db 290 IQTYQTAKQSGYTNFNFSFLSSFYKESNFM-----YGSYHPSCKF-----331

Qy 187 YHYVDRKLDVTNHNKDYTSVQRETSIARKPL-----VACCF 223
Db 332 -----RLETINGLWENSL---SVSIAYGLQGGCKQSVFKGRATCCY 371

RESULT 7
YDEM_ECOLI
ID_YDEM_ECOLI STANDARD; PRT; 385 AA.
AC P76134; P77755;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydem.
GN YDEM OR B1497.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
```

```
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RP [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isoo K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Mishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: BELONGS TO THE ASLB/ATSB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000247; AAC74570.1; ALT_INIT.
CC DR EMBL; D90791; BAA15168.1; -.
CC DR EMBL; D90792; BAA15171.1; -.
CC DR EcoGene; EG13795; ydem.
CC DR InterPro; IPR007197; Radical_SAM.
CC DR Pfam; PF04055; Radical_SAM_1.
CC DR Hypothetical protein; Complete proteome.
CC SEQUENCE 385 AA; 44518 MW; 964E34F73E680329 CRC64;

Query Match 6.6%; Score 83.5; DB 1; Length 385;
Best Local Similarity 19.6%; Pred. No. 3.9;
Matches 55; Conservative 46; Mismatches 104; Indels 75; Gaps 12;

Qy 10 KVMNSGTVNGHYFVEVGGKGP-----YEGQTVRLAV-TKGGLPFF 51
Db 44 KQYIAASGNQVYFTWQG---GEPTLAGLDFFKVIHYQRYAGOKRIFNALQTNILLNN 100

Qy 52 AWDILSPQCQVGSIPFTKYPEDIDYVKQSPGGRYTW-----ERIMNPE-DGAVCTVS 103
Db 101 EWCAFLKEHEFLVIGISIDGPOELHRYRRSSNGTFAKVIATAIERLKSQYVEFNTLTVI 160

Qy 104 NDSSIQGNCFIYH-----VKFSG-----LNFPN--GP 129
Db 161 NVNVNHYPLEVYHFKLSIGSKHMQFIELTGTNIDFSGHSENTFRIDFSVPETAYGK 220

Qy 130 VMQKKTQGWEPN-TERLFARDGMLIGNFMALKLEGGHYLCFEFKSYTKARKPVKMPG---186
Db 221 FMSTIFMQWVKNDVGEIFRQFESFVSRL-----GNHTSCIFQESCKNLVVESNGDI 275

Qy 187 ---YHYVDRKLDVTNHNK-DYTSVQRETSIARKPLVACC 222
Db 276 YECDFVTPQYKIGNINKSELKTWNVSOLTAQKRIPAKC 315

RESULT 8
RELN_MOUSE
ID_RELN_MOUSE STANDARD; PRT; 3461 AA.
AC Q60841; Q9CUGA6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Reelin precursor (EC 3.4.21.-) (reeler protein).
GN RELN OR RL.
```


OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Cerebellum;
RX MEDLINE=95231649; PubMed=7715726;
RA D'Arcangelo G., Miao G.G., Chen S.-C., Soares H.D., Morgan J.I.,
RA Curran T.;
RT "A protein related to extracellular matrix proteins deleted in the
RT mouse mutant reeler.";
RL Nature 374:719-723(1995).
RN [2]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=98086481; PubMed=9417311;
RA Royaux I., Lambert de Rouvroit C., D'Arcangelo G., Demirov D.,
RA Goffinet A.M.;
RT "Genomic organization of the mouse reeler gene.";
RL Genomics 46:240-250(1997).
RN [3]
RP SEQUENCE OF 2152-3461 FROM N.A. (ISOFORM 1).
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=95375789; PubMed=7647795;
RA Hirotsune S., Takahara T., Sasaki N., Hirose K., Yoshiki A.,
RA Ohashi T., Kusakabe M., Murakami Y., Muramatsu M., Watanabe S.,
RA Nakao K., Katsuki M., Hayashizaki Y.;
RT "The reeler gene encodes a protein with an EGF-like motif expressed by
RT pioneer neurons.";
RL Nat. Genet. 10:77-83(1995).
RN [4]
RP SEQUENCE OF 3044-3461 FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97141547; PubMed=8987733;
RA D'Arcangelo G., Nakajima K., Miyata T., Ogawa M., Mikoshiba K.,
RA Curran T.;
RT "Reelin is a secreted glycoprotein recognized by the CR-50 monoclonal
RT antibody.";
RL J. Neurosci. 17:23-31(1997).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=21634904; PubMed=11699558;
RA Quattrocchi C.C., Wannenres F., Persico A.M., Ciafre S.A.,
RA D'Arcangelo G., Farace M.G., Keller F.;
RT "Reelin is a serine protease of the extracellular matrix.";
RL J. Biol. Chem. 277:303-309(2002).
RN [7]
RP TISSUE SPECIFICITY.
RX MEDLINE=97325946; PubMed=9182958;
RA Schiffmann S.N., Bernier B., Goffinet A.M.;
RT "Reelin mRNA expression during mouse brain development.";
RL Eur. J. Neurosci. 9:1055-1071(1997).
RN [8]
RP ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RX MEDLINE=99263436; PubMed=10328932;
RA Lambert de Rouvroit C., Bernier B., Royaux I., de Bergcyck V.,
RA Goffinet A.M.;
RT "Evolutionarily conserved, alternative splicing of reelin during brain
RT development.";
RL Exp. Neurol. 156:229-238(1999).
RN [9]
RP BINDING TO VLDLR AND APOER2.
RX MEDLINE=20036019; PubMed=10571241;
RA Hiesberger T., Trommsdorff M., Howell B.W., Goffinet A.M., Mumby M.C.,
RA Cooper J.A., Herz J.;
RT "Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces
RT tyrosine phosphorylation of disabled-1 and modulates tau
RT phosphorylation.";
RL Neuron 24:481-489(1999).
RN [10]
RP FUNCTION.
RX MEDLINE=20359755; PubMed=10880573;
RA Yip J.W., Yip Y.P.L., Nakajima K., Capriotti C.;
RT "Reelin controls position of autonomic neurons in the spinal cord.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:8612-8616(2000).
CC -!- FUNCTION: Extracellular matrix serine protease that plays a role
CC in layering of neurons in the cerebral cortex and cerebellum.
CC Regulates microtubule function in neurons and neuronal migration.
CC Affects migration of sympathetic preganglionic neurons in the
CC spinal cord, where it seems to act as a barrier to neuronal
CC migration. Enzymatic activity is important for the modulation of
CC cell adhesion. Binding to the extracellular domains of lipoprotein
CC receptors VLDLR and ApoER2 induces tyrosine phosphorylation of
CC Dab1 and modulation of Tau phosphorylation.
CC -!- SUBUNIT: Binds to the ectodomains of VLDLR and ApoER2.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q60841-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q60841-2; Sequence=VSP_005577;
CC Name=3;
CC IsoId=Q60841-3; Sequence=VSP_005578;
CC -!- TISSUE SPECIFICITY: The major isoform 1 is neuron-specific. It is
CC abundantly produced during brain ontogenesis by the Cajal-Retzius
CC cells and other pioneer neurons located in the telencephalic
CC marginal zone and by granule cells of the external granular layer
CC of the cerebellum. Expression is located in deeper layers in the
CC developing hippocampus and olfactory bulb, low levels of
CC expression are also detected in the immature striatum. At early
CC developmental stages, expressed also in hypothalamic
CC differentiation fields, tectum and spinal cord. A moderate to low
CC level of expression occurs in the septal area, striatal fields,
CC habenular nuclei, some thalamic nuclei, particularly the lateral
CC geniculate, the retina and some nuclei of the reticular formation
CC in the central field of the medulla. Very low levels found in
CC liver and kidney. No expression in radial glial cells, cortical
CC plate, Purkinje cells and inferior olivary neurons. The minor
CC isoform 2 is only expressed in non neuronal cells. The minor
CC isoform 3 is found in the same cells as isoform 1, but is almost
CC undetectable in retina and brain stem.
CC -!- DEVELOPMENTAL STAGE: First detected at embryonic day 11.5.
CC Expression increases up to birth and remains high from post-natal
CC day 2 to 11 in both cerebellum and fore/midbrain. Expression
CC declines thereafter and is largely brain specific in the adult.
CC -!- DOMAIN: The basic C-terminal region is essential for secretion.
CC -!- PTM: N-glycosylated and to a lesser extent also O-glycosylated.
CC -!- DISEASE: Defects in reelin are the cause of the autosomal recessive
CC reeler (rl) phenotype which is characterized by impaired motor
CC coordination, tremors and ataxia. Neurons in affected mice fail to
CC reach their correct locations in the developing brain, disrupting
CC the organization of the cerebellar and cerebral cortices and other

laminated regions.
 CC -!- SIMILARITY: Belongs to the reelin family.
 CC -!- SIMILARITY: Contains 8 EGF-like domains.
 CC -!- SIMILARITY: Contains 15 BNR repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U24703; BAB91599.1; --
 DR EMBL; D63520; BAA09788.1; ALT INIT.
 DR EMBL; AK017094; BAB30592.1; --
 DR MGd; MGI:103022; ReIn.
 DR GO; GO:0005615; C:extracellular space; IDA.
 DR GO; GO:0007420; P:brain development; IMP.
 DR GO; GO:0016477; P:cell migration; IMP.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR002860; GH BNR.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002861; Reeler.
 DR Pfam; PF02012; BNR; 15.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF02014; Reeler; 1.
 DR SMART; SM00181; EGF; 5.
 DR PROSITE; PS00022; EGF 1; 7.
 DR PROSITE; PS01186; EGF 2; 6.
 DR PROSITE; PS00026; EGF 3; 5.
 KW Hydrolyase; Serine protease; Developmental protein; Matrix protein;
 KW Cell adhesion; EGF-like domain; Glycoprotein; Repeat; Signal;
 KW Alternative splicing.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 3461 REELIN.
 FT DOMAIN 40 172 REELIN.
 FT DOMAIN 671 702 EGF-LIKE 1.
 FT DOMAIN 1030 1061 EGF-LIKE 2.
 FT DOMAIN 1409 1442 EGF-LIKE 3.
 FT DOMAIN 1765 1796 EGF-LIKE 4.
 FT DOMAIN 2129 2161 EGF-LIKE 5.
 FT DOMAIN 2478 2509 EGF-LIKE 6.
 FT DOMAIN 2853 2884 EGF-LIKE 7.
 FT DOMAIN 3228 3260 EGF-LIKE 8.
 FT REPEAT 593 604 BNR 1.
 FT REPEAT 799 810 BNR 2.
 Query Match 6.6%; Score 83.5; DB 1; Length 3461;
 Best Local Similarity 20.3%; Pred. No. 53;
 Matches 57; Conservative 24; Mismatches 75; Indels 125; Gaps 12;
 QY 12 YMSGTVNGHYFEVEGDGK---GKPYEGEQTIVRLAVTKG---GGLPFANDI---LSPQ 59
 DB 2080 YYAGTTQGWREVVHFKLHLCG-----SVRFRTYGGYPAGSQPVTHVDNVIQGP 2132
 QY 60 QQ---YG-----SIPFTKYPEDIPYVVKQSPFGRYTWERIMNF----- 94
 DB 2133 CEEMCYGHGSCINGTKCICDPGYSGPCKLSTKNPDLKDDFGQLSDRLLMSGGKPS 2192
 QY 95 -----EDGAVCTVNSDSSIQGNCFIYHVKF-----SGLNFPNGP 129
 DB 2193 RKGILLSSGNLFFNEDGLRMLVTRDLS-----HARFVQFFMRLLGCGKGVDPDRSQP 2246
 QY 130 VM-----QKKTQGWENTERLF--- 146
 DB 2247 VLLOYSINGLSWLLQELFNSNSVGRVIAEMPLKARGSTRLRWQPSENGHYSP 2306
 QY 147 -ARDGMLIGNNFMALKLEGGHYLCEPKSTYKARKPKVMPG 186
 DB 2307 WVIDQILIGNI-----SGNTVLEDDFSLDSRKLHLHPG 2341

RESULT 9
 SYA STAA
 ID SYA STAA STANDARD; PRT; 876 AA.
 AC Q99TNI;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (Alars).
 GN ALAS OR SAV1618 OR SA1446.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878, 158879;
 RN [1]
 RZ SEQUENCE FROM N.A.
 RP STRAIN=Mu50 / ATCC 700699, and N315;
 RC MEDLINE=21311952; PubMed=1418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RL Lancet 357:1225-1240(2001).
 CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
 CC diphosphate + L-alanyl-tRNA(Ala).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP003362; BAB57780.1; --
 DR EMBL; AP003134; BAB42710.1; --
 DR PIR; A89944; A89944.
 DR HAMAP; MF 00036; -. 1.
 DR InterPro; IPR003156; DHHA1.
 DR InterPro; IPR002318; tRNA-synt 2c.
 DR InterPro; IPR006193; tRNA_synt_Ala.
 DR Pfam; PF02272; DHHA1; 1.
 DR Pfam; PF01411; tRNA-synt 2c; 1.
 DR PRINTS; PR00980; TRNASYNTHALA.
 DR TIGRFAMS; TIGR00344; alaS; 1.
 DR PROSITE; PS50860; AA TRNA LIGASE II ALA; 1.
 KW Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 SQ SEQUENCE 876 AA; 98538 MW; 2B2BC79041AC264F CRC64;
 Query Match 6.5%; Score 82.5; DB 1; Length 876;
 Best Local Similarity 21.9%; Pred. No. 13;
 Matches 46; Conservative 25; Mismatches 60; Indels 79; Gaps 11;
 QY 16 TVNCH-YFEVEGDGKPYEGEQTIVRLAVTKGGLPFANDILSPQCGYSGIP----FTKY 70
 DB 81 TARHHTFEFELNFSIGDYFKQEAIE-----FAWEFLTSKWMGMEPDKLYVTH 130
 QY 71 PEDIPDYVVKQSPFGRYTWERIMNPFEDGAVCTVNSDSSIQGNCFIYHVKFSGLNPPNGPV 130
 DB 131 PEDMEAY-----NTHKDIGLESRII-----RIEGN-----FWDIGGPGSG-- 167
 QY 131 MQKKTQGWENTERLFPARDGMLIGNNFMALKLEGGHYLCEPKSTYKARKPKVMPGHHYV 190
 DB 168 -----ENTETFYDTR-GEAYGQDDPABEMYPGGEN----- 195
 QY 191 DRKLDV-----TNNKDD--YTSVEQREI 211

FT CARBOHYD 3186 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3413 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3440 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 3431 Missing (in isoform 2).
 FT VARSPLIC 3430 /FTID=VSP 005579.
 FT VARSPLIC 3462 Missing (in isoform 3).
 FT CONFLICT 336 H -> R (IN REF. 2).
 FT CONFLICT 2714 V -> L (IN REF. 2).
 SQ SEQUENCE 3462 AA; 387525 MW; FCC898B090E035F6 CRC64;
 Query Match 6.4%; Score 81.5; DB 1; Length 3462;
 Best Local Similarity 19.9%; Pred. No. 80;
 Matches 56; Conservative 25; Mismatches 75; Indels 125; Gaps 12;
 QY 12 YMSGTVNGHYEVEGDGK-----GKPYEGEQTIVRLAVTKG-----GFLPFWNDLSPQ 59
 Db 2081 YAGTITQWRREVHFKLHLCG-----SVRFYQGFYAGPAGQPVTAIDNVYIGPQ 2133
 QY 60 CQ-----YCSI-----PFTKYPEDIPDYVKQSPGRTYTWERIMNF----- 94
 Db 2134 CEEMCCGHGSCVNGTKICIDPGYSGTCKISTKNDFDLKDFEGQLESDFRLMSGGKPS 2193
 QY 95 -----EDGAVCTVSNDSIQGNCFTYHVKF-----SGINFFPNXP 129
 Db 2194 RKCGLTSSGNLFFNEDGLMLVTRDLDS-----HARFVQFFWRLGCGKGVDPDRSQP 2247
 QY 130 VM-----OKTQWEPNTERLF--- 146
 Db 2248 VLLQYSLNGLSWSLQBLFLFSNSNVGRIYALEMPLKARSGSTLRWQPSENGHFYSP 2307
 QY 147 -ARDQMLIGNFMALKLEGGHYLCFEKSTYKARKPKVMPG 186
 Db 2308 WVIDQLIGNI-----SGNTVLEDDFSTLDRKWLHPG 2342

RESULT 13
 ENAM HUMAN
 ID ENAM HUMAN STANDARD; PRT; 1142 AA.
 AC Q9NEM1_Q9H3D1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Enamelin precursor.
 GN ENAM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hu C.-C., Qian Q., Zhang C., Fukae M., Uchida T., Simmer J.P.;
 RT "cDNA sequence of human enamel.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1014-1142 FROM N.A.
 RA MEDLINE=20489450; PubMed=11037750;
 RX Dong J., Gu T.T., Simmons D., MacDougall M.;
 RT "Enamelin maps to human chromosome 4q21 within the autosomal dominant
 amelogenesis imperfecta locus.";
 RL Eur. J. Oral Sci. 108:353-358(2000).
 CC -!- FUNCTION: Involved in the mineralization and structural
 organization of enamel. Involved in the extension of enamel during
 the secretory stage of dental enamel formation.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- TISSUE SPECIFICITY: Expressed in tooth particularly in
 odontoblast, ameloblast and cementoblast.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF125373; AAG43242.1; -.
 DR EMBL; AF210247; AAF73847.1; -.
 DR EMBL; HGNC:3344; ENAM.
 DR MIM; 606585; -.
 DR GO; GO:0005578; C:extracellular matrix; NAS.
 DR GO; GO:0030345; F:structural constituent of tooth enamel; NAS.
 DR GO; GO:0030282; P:bone mineralization; NAS.
 DR GO; GO:0042476; P:odontogenesis; NAS.
 KW Biomineralization; Extracellular matrix; Glycoprotein; Signal.
 FT SIGNAL 1 39 POTENTIAL.
 FT CHAIN 40 1142 ENAMELIN.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 934 934 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1142 AA; 128745 MW; 77419C4375EAD6EC CRC64;
 Query Match 6.4%; Score 81; DB 1; Length 1142;
 Best Local Similarity 29.5%; Pred. No. 24;
 Matches 23; Conservative 12; Mismatches 35; Indels 8; Gaps 3;
 QY 18 NGHFEVGEKGGKPYEGEQTIVRLAVTKG-----GFLPFWNDLSPQCYGSIPTKYPE 72
 Db 493 NSYI--PRGDSRKVPNSDGTQSQNLPGKIVLGRMPYESETNQSELKHSYQPAVTYPE 550
 QY 73 DIPYVKQSPF-GRYTWE 89
 Db 551 EIPSPAKEHFFPAGRNTWD 568
 RESULT 14
 ID ATAI CHICK STANDARD; PRT; 994 AA.
 AC P13585;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sarcoplasmic/endoplasmic reticulum calcium ATPase 1 (BC 3.6.3.8)
 DE (Calcium pump 1) (SERCA1) (SR Ca(2+)-ATPase 1) (Calcium-transporting
 ATPase sarcoplasmic reticulum type, fast twitch skeletal muscle
 isoform) (Endoplasmic reticulum class 1/2 Ca(2+) ATPase).
 GN ATP2A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89313743; PubMed=2526293;
 RA Karin N.J., Kaprielian Z., Fambrough D.M.;
 RT "Expression of avian Ca2+-ATPase in cultured mouse myogenic cells.";
 RL Mol. Cell. Biol. 9:1978-1986(1989).
 CC [2]
 RP REVISIONS.
 RA Karin N.J.;
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: This magnesium-dependent enzyme catalyzes the hydrolysis
 of ATP coupled with the translocation of calcium from the cytosol
 to the sarcoplasmic reticulum lumen. Contributes to calcium
 sequestration involved in muscular excitation/contraction (By
 similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Ca(2+) (Cis) = ADP + phosphate +
 Ca(2+) (Trans).

This page blank (uspto)

Result No.	Score	Query %		Length	DB	ID	Description
		Match	Length				
1	1138	89.7	221	5	Q95P04	Q95P04	goniopora t
2	780	61.5	225	5	Q9U6Y8	Q9U6Y8	discosoma s
3	750	59.1	230	5	Q9GTJ7	Q9GTJ7	discosoma s
4	737.5	58.2	232	5	Q9U6Y7	Q9U6Y7	discosoma s
5	704.5	55.6	236	5	Q8T6U0	Q8T6U0	dendronephth
6	703	55.4	225	5	Q963F5	Q963F5	montastraea
7	698	55.0	225	5	Q7Z0W4	Q7Z0W4	montastraea
8	672	53.0	225	5	Q7Z0W5	Q7Z0W5	montastraea
9	670	52.8	266	5	Q9U6Y3	Q9U6Y3	clavularia
10	667.5	52.6	227	5	Q7Z0W6	Q7Z0W6	montastraea
11	665.5	52.5	225	5	Q7Z0W9	Q7Z0W9	montastraea
12	665.5	52.5	227	5	Q962P9	Q962P9	montastraea
13	665.5	52.5	227	5	Q7Z0W8	Q7Z0W8	montastraea
14	664	52.4	225	5	Q95UA7	Q95UA7	montastraea
15	655.5	51.7	234	5	Q8U5F2	Q8U5F2	montastraea
16	654.5	51.6	225	5	Q8U5F1	Q8U5F1	montastraea

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```
QY 121 GLNPPNGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 180
Db 122 GLNFPNGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKAKK 181
QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQREISIAKRLVA 220
Db 182 PVKMPGHHYVDRKLDVTNNHKDYSVEQREISIAKRPVA 221

RESULT 2
Q9U6Y8 PRELIMINARY; PRT; 225 AA.
AC Q9U6Y8
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP583.
OS Discosoma sp.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=86600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zairaisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RA "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zairaisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF168419; AAF03369.1; -.
DR PDB; 1G7K; 07-NOV-01.
DR PDB; 1GGX; 06-DEC-00.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green.fl_protein; 1.
SQ SEQUENCE 225 AA; 25931 MW; FBF9A5369778F689 CRC64;

Query Match 61.5%; Score 780; DB 5; Length 225;
Best Local Similarity 62.6%; Pred. No. 5.3e-66;
Matches 144; Conservative 28; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAKMTKYVMSGTNGHYFVEGDKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 6 NVIKEMFRKVRMEGTNGHFEIEGEGRPYEGHNTVKLVTKGKPLPFAWDILSPQF 65
QY 61 QYGISIFTKYKPEDIPYVVKSPGRTYWRIMNFEDGAVCTVSNDSIQNCFIYHVKS 120
Db 66 QYGSKYVVKHPADIPYKLSPEFGKRWVNFEDGGVTVTQDSSLDQGCFFIYKVKFI 125
QY 121 GLNFPNGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 180
Db 126 GVNFPDGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 185
QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQREISIAKRLVA 220
Db 186 PVQLPFGYVYVDRKLDVTNNHKDYSVEQREISIAKRLVA 221

RESULT 3
Q9GTU7 PRELIMINARY; PRT; 230 AA.
AC Q9GTU7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
QY 121 GLNFPNGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 180
Db 126 GVNFPDGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 185
QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQREISIAKRLVA 220
Db 186 PVQLPFGYVYVDRKLDVTNNHKDYSVEQREISIAKRLVA 221

RESULT 4
Q9U6Y7 PRELIMINARY; PRT; 232 AA.
AC Q9U6Y7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP483.
OS Discosoma striata.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=105400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zairaisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RA "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
DR EMBL; AF168420; AAF03370.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green.fl_protein; 1.
SQ SEQUENCE 232 AA; 26435 MW; AA8F18EE283CE4D CRC64;
```

```
DE Red fluorescent protein.
GN FP593.
OS Discosoma sp. SSAL-2000.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=137428;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20434599; PubMed=10981720;
RA Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA Lukyanov S.A.;
RT "Novel fluorescent protein from Discosoma coral and its mutants
RT possesses a unique far-red fluorescence.";
RL FEBS Lett. 479:127-130(2000).
DR EMBL; AF272711; AAG16224.1; -.
DR HSSP; P42212; 1BFP.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green.fl_protein; 1.
SQ SEQUENCE 230 AA; 26370 MW; 5215B1B436D67B51 CRC64;

Query Match 59.1%; Score 750; DB 5; Length 230;
Best Local Similarity 62.5%; Pred. No. 3.8e-63;
Matches 140; Conservative 36; Mismatches 42; Indels 6; Gaps 2;

QY 1 SVIAKMTKYVMSGTNGHYFVEGDKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 6 NVIKEMFRKVRMEGTNGHFEIEGEGRPYEGHNTVKLVTKGKPLPFAWDILSPQF 65
QY 61 QYGISIFTKYKPEDIPYVVKSPGRTYWRIMNFEDGAVCTVSNDSIQNCFIYHVKS 120
Db 66 QYGSKYVVKHPADIPYKLSPEFGKRWVNFEDGGVTVTQDSSLDQGCFFIYKVKFI 125
QY 121 GLNFPNGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 180
Db 126 GVNFPDGPVQKKTQGWENTRLFARDGMLIGNFMALKLEGGHYLCFEKSTYKARK 185
QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQREISIAKRLVA 220
Db 186 PVQLPFGYVYVDRKLDVTNNHKDYSVEQREISIAKRLVA 221

RESULT 4
Q9U6Y7 PRELIMINARY; PRT; 232 AA.
AC Q9U6Y7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP483.
OS Discosoma striata.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=105400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zairaisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RA "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
DR EMBL; AF168420; AAF03370.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green.fl_protein; 1.
SQ SEQUENCE 232 AA; 26435 MW; AA8F18EE283CE4D CRC64;
```

Query Match 58.2%; Score 737.5; DB 5; Length 232;
Best Local Similarity 59.7%; Pred. No. 6e-62;
Matches 129; Conservative 39; Mismatches 47; Indels 1; Gaps 1;

QY 1 SVIAKOMTYKVMSTGVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
DB 6 SVIKEMILDIHLEGTFFNGHYFEIKGKGQPNEGTNTVTLVTKGGPLPFGWHILCPQF 65
QY 61 QYGSIPFTKYPEDIPDYVKQSPGPGRYTWERIMNFDGAVCTVSDSSIQGNCFIYHVKFS 120
DB 66 QYGNKAFVHPEDNIHDLKLSFPPEGYTWERSMHFEDGGLCCIINDISLTGNCFYDIKFT 125
QY 121 GLNFPNGPVMOKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKARK 180
DB 126 GLNFPNGPVMOKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKARK 185
QY 181 -PVKMPGHHYVDRKLDVTHNKNKYTSVEQREISAR 215
DB 186 AALKRMPGHYVDTKLVWINDKFNKVEHEIAVAR 221

RESULT 5
Q8TG6UO PRELIMINARY; PRT; 236 AA.

ID Q8TG6UO AC
Q8TG6UO DT
01-JUN-2002 (TREMBLrel. 21, Created)
01-JUN-2002 (TREMBLrel. 21, Last sequence update)
01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Dendronephthya sp. SBAU-2002.
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC Nephthelidae; Dendronephthya.
ON NCBI_TaxID=191210;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21927629; PubMed=11929996;
RA Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.,
RT "Diversity and evolution of the green fluorescent protein family.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR EMBL; AF420591; AAM10625.1; --
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000917; GFP_like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 236 AA; 26840 MW; CE1707CFF9334A90 CRC64;

Query Match 55.6%; Score 704.5; DB 5; Length 236;
Best Local Similarity 55.3%; Pred. No. 8.4e-59;
Matches 121; Conservative 46; Mismatches 51; Indels 1; Gaps 1;

QY 1 SVIAKOMTYKVMSTGVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
DB 2 NLIKEDMRVKHMEGNVANGHAFVIEGEGKGRPYEGTQTLNLTVEGAPLPFSYDILATL 61
QY 61 QYGSIPFTKYPEDIPDYVKQSPGPGRYTWERIMNFDGAVCTVSDSSIQGNCFIYHVKFS 120
DB 62 HYGNRVFTPEADITDYFKQSPFGYSWERNYEDKIGICTIRSDIEGDCFFQNTFRN 121
QY 121 GLNFPNGPVMOKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKARK 180
DB 122 GNNFPNGPVMOKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKARK 181
QY 181 PVKMPGHHYVDRKLDVTHNKNKYTSVEQREISAR-KPL 218
DB 182 VVQLPDYHFDVDRHRIETLSNDSYNNKVKLYEHGVARYSPL 220

RESULT 6
Q963F5 PRELIMINARY; PRT; 225 AA.

ID Q963F5 AC
Q963F5 DT
01-DEC-2001 (TREMBLrel. 19, Created)
01-DEC-2001 (TREMBLrel. 19, Last sequence update)
01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
ON NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=mc6;
RA Kelman I.V., Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great
Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181557; AAO61603.1; --
SQ SEQUENCE 225 AA; 25827 MW; A600ADD716C5921E CRC64;

Query Match 55.0%; Score 698; DB 5; Length 225;
Best Local Similarity 55.8%; Pred. No. 3.3e-58;
Matches 120; Conservative 43; Mismatches 52; Indels 0; Gaps 0;

QY 1 SVIAKOMTYKVMSTGVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
DB 1 SVIAKOMTYKVMSTGVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60

```
Db 2 SVIKPDMKIKLRMEGAVNGHNFVIEGEGKPGEGTQTLNLTVEGGGLPFAVDILTAAF 61
QY 61 QYGSIPPTKYPEDIPDYVKQSPFGRYTWERIMNPFEDGAVCTVNSDSSIOGNCFIYHVKFS 120
Db 62 QYGNRAFTKYPDIADYFKQSPFEGYSWERSMTYEDQGICILKSDIRWEGDCFIYEIRYD 121
QY 121 GLNFPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGHYLCFEKFSYKARK 180
Db 122 GVNFPSPGPMQKTKLWEPSTEMKYVYRDGVILKGDVNMALLEGGHYRCDFRSTYKAKK 181
QY 181 PVKMPGHHYVDRKLDVTNNHNDYTSVEQREISAR 215
Db 182 RVQLPDYHFVDRHRIELSHDNDYNTVLSNAEAR 216
```

```
RESULT 8
Q720W5
ID Q720W5 PRELIMINARY; PRT; 225 AA.
AC Q720W5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein.
OS Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc5;
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V., Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181556; AAO61602.1; -.
SQ SEQUENCE 225 AA; 25843 MW; 13708587B7D93E35 CRC64;
```

```
Query Match 53.0%; Score 672; DB 5; Length 225;
Best Local Similarity 53.4%; Pred. No. 9.7e-56;
Matches 117; Conservative 45; Mismatches 57; Indels 0; Gaps 0;

QY 1 SVIAKQMTYKVMGTVNGHYFEVGEKGKPYEGEQTVRVLAVTKGGPLPFAWDILSPQC 60
Db 2 SVIKSVMKIKLHMDGIVNGHKFMTIGEGKPGEGTHTIILKVKEGGGLPFAVDILTAFF 61
QY 61 QYGSIPPTKYPEDIPDYVKQSPFGRYTWERIMNPFEDGAVCTVNSDSSIOGNCFIYHVKFS 120
Db 62 QYGNRVFTKYPKDIIPDYFKQSPFEGYSWERSMTYEDQGVCTVTSIDIKLEGDCFFYEIRFY 121
QY 121 GLNFPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGHYLCFEKFSYKARK 180
Db 122 GVNFPSPGPMQKTKLWEPSTENMYVYRDGVLLGDVTSVTLLEGDKHRCNFRSTYKAKK 181
QY 181 PVKMPGHHYVDRKLDVTNNHNDYTSVEQREISARPLV 219
Db 182 GVVLPDYHFVDRHRIELSHDNDYNTVEYENAVARPSML 220
```

```
RESULT 9
Q906Y3
ID Q906Y3 PRELIMINARY; PRT; 266 AA.
AC Q906Y3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP484.
OS Clavularia sp.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC Clavulariidae; Clavularia.
OX NCBI_TaxID=86521;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Pradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
DR EMBL; AF168424; AAF03374.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green fl protein; 1.
SQ SEQUENCE 266 AA; 30450 MW; B4E97406E2708854 CRC64;
```

```
Query Match 52.8%; Score 670; DB 5; Length 266;
Best Local Similarity 54.6%; Pred. No. 1.8e-55;
Matches 119; Conservative 39; Mismatches 60; Indels 0; Gaps 0;

QY 2 VIAKQMTYKVMGTVNGHYFEVGEKGKPYEGEQTVRVLAVTKGGPLPFAWDILSPQC 61
Db 45 VIKPDMKIKLMEGTVNGHAFVIEGEGKPYDGTHTLNLVKGAGPLPFSYDILSNAFQ 104
QY 62 YGSIPPTKYPEDIPDYVKQSPFGRYTWERIMNPFEDGAVCTVNSDSSIOGNCFIYHVKFS 121
Db 105 YGNRALTKYPDDIADYFKQSPFEGYSWERTMTPTEDKGIVKVKSDISMEEDSFYIEIRFDG 164
QY 122 LNFPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGHYLCFEKFSYKARKP 181
Db 165 MNFPNGPVMQKTKLWEPSTENMYVYRDGVLLGDVTSVTLLEGDKHRCNFRSTYKAKKV 224
QY 182 VKMPGHHYVDRKLDVTNNHNDYTSVEQREISARPLV 219
Db 225 VKLPDYHFVDRHRIELSHDNDYNTVLYENAVARYSL 262
```

```
RESULT 10
Q720W6
ID Q720W6 PRELIMINARY; PRT; 227 AA.
AC Q720W6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc4;
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V., Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181555; AAO61601.1; -.
SQ SEQUENCE 227 AA; 26055 MW; 4BE2CB64FDB0E890 CRC64;
```

```
Query Match 52.6%; Score 667.5; DB 5; Length 227;
Best Local Similarity 53.4%; Pred. No. 2.6e-55;
Matches 119; Conservative 43; Mismatches 57; Indels 3; Gaps 1;

QY 1 SVIAKQMTYKVMGTVNGHYFEVGEKGKPYEGEQTVRVLAVTKGGPLPFAWDILSPQC 60
Db 2 SVIKPDMKIKLRMEGAVNGHNFVIEGEGKPGEGTQTLNLTVEGGGLPFAVDILTAFF 61
QY 61 QYGSIPPTKYPEDIPDYVKQSPFGRYTWERIMNPFEDGAVCTVNSDSSIOGNCFIYHV 117
Db 62 DYGNRVFAKYPKDIIPDYFKQSPFEGYSWERSMTYEDQGIATNDITMMKGVDDCFLYKI 121
QY 118 KFSGLNFPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGHYLCFEKFSYK 177
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:19:43 ; Search time 486.141 Seconds
(without alignments)
149.169 Million cell updates/sec

Title: US-09-890-463-3
Perfect score: 1268
Sequence: 1 SVIAKQMTYKVMSTGVNGH.....SIARKPLVACFFRVKSRHK 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	785	61.9	225	15	US-10-442-148A-7
2	785	61.9	239	15	US-10-442-148A-8
3	782	61.7	225	14	US-10-315-920-6
4	781	61.6	225	14	US-10-121-258-20
5	781	61.6	225	14	US-10-315-920-4
6	781	61.6	236	16	US-10-314-936-2
7	781	61.6	236	16	US-10-314-936-4
8	780	61.5	225	9	US-09-999-745-67
9	780	61.5	225	10	US-09-866-538-12
10	780	61.5	225	10	US-09-794-308-12
11	780	61.5	225	10	US-09-865-291-12
12	780	61.5	225	12	US-10-132-067-4
13	780	61.5	225	13	US-10-006-922-12
14	780	61.5	225	14	US-10-081-864-8
15	780	61.5	225	14	US-10-121-258-1

16	780	61.5	225	14	US-10-315-920-2
17	780	61.5	225	15	US-10-370-570-56
18	780	61.5	225	15	US-10-406-618-32
19	780	61.5	225	16	US-10-433-640-13
20	780	61.5	240	14	US-10-152-296-2
21	780	61.5	240	16	US-10-739-656-2
22	780	61.5	487	15	US-10-343-977-1
23	780	61.5	506	15	US-10-343-977-2
24	780	61.5	545	14	US-10-214-932-52
25	780	61.5	547	15	US-10-343-977-3
26	780	61.5	548	14	US-10-214-932-76
27	775	61.1	225	16	US-10-423-688A-41
28	773	61.0	227	13	US-10-006-922-10
29	767	60.5	226	14	US-10-121-258-6
30	766	60.4	225	14	US-10-121-258-24
31	765	60.3	225	13	US-10-006-922-44
32	765	60.3	225	14	US-10-081-864-12
33	765	60.3	225	14	US-10-121-258-4
34	761.5	60.1	226	16	US-10-724-178-12
35	750	59.1	230	13	US-10-006-922-18
36	750	59.1	230	14	US-10-161-403-40
37	737.5	58.2	232	14	US-10-133-973-5
38	737.5	58.2	232	15	US-10-370-570-64
39	729.5	57.5	232	16	US-10-724-178-11
40	714	56.3	225	14	US-10-121-258-8
41	714	56.3	225	16	US-10-724-178-16
42	703	55.4	225	15	US-10-370-570-61
43	687	54.2	205	13	US-10-006-922-46
44	670	52.8	231	14	US-10-133-973-6
45	670	52.8	266	13	US-10-006-922-4

ALIGNMENTS

RESULT 1
US-10-442-148A-7
; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: HIROTA, KIYONORI
; APPLICANT: IWAKURA, MASAHITO
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583.0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

Query Match	61.9%;	Score 785;	DB 15;	Length 225;
Best Local Similarity	62.6%;	Pred. No. 1.3e-76;		
Matches 144;	Conservative 28;	Mismatches 44;	Indels 14;	Gaps 1;
QY	1	SVIAKQMTYKVMSTGVNGHYFEVGDGKPKYEGEQTVRLATYKGGPLPFAWDILSPQC	60	
Db	6	NVITEFMKPKVMKEGTVNGHEFEIEGEGRPEYGHNTVKLVKTKGGPLPFAWDILSPQF	65	
QY	61	OYGSIPFFKYPIEDIDYVKQSPRGYTWERINNEEDGAVCTVSDSSIOGNCFIYHVKFS	120	
Db	66	OYGSKYVYKHPADIDYKRLSPFEKGRVNMVNFVGGVATVTQDSSLODGGCFIYKVKFI	125	
QY	121	GLNEFPNGVPVMOKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHYLCEPKSTYKARK	180	


```
QY 121 GLNFPNGPVMOCKTQGWEPNTERLFARDGMLIGNFMALKLEGGHYLCBFKSTYKARK 180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 GVNFDGPFVMOCKTMGWEASTERLYPRDGVLGKGIHAKLKLDGSHYLVFVKSIYMAKK 185
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 PVKMPGYHYVDKRLDVTNNKDYTSVEQREISIAKRLPVACCFRVRKSRH 230
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 PVQLPGYHYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221

RESULT 12
US-10-132-067-4
; Sequence 4, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
; FILE REFERENCE: 021362-000600US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; OTHER INFORMATION: red fluorescent protein (dsred)
US-10-132-067-4

Query Match 61.5%; Score 780; DB 12; Length 225;
Best Local Similarity 62.6%; Pred. No. 4.4e-76;
Matches 144; Conservative 28; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAQMTYKYVMGTVNGHYFEVGDGKGKPYEGEQTURLAVTTKGGPLPFAWDILSPQC 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 6 NVIKFMRPKVRMEGTVNGHFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSPQF 65
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 QXGSIPTFKYPEDIPDYVKQSPFGRYTWERIMNFEDGAVCTVSNDSIIQNCFCFIYHKFS 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 66 QXGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFIYKVKFI 125
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 GLNFPNGPVMOCKTQGWEPNTERLFARDGMLIGNFMALKLEGGHYLCBFKSTYKARK 180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 GVNFDGPFVMOCKTMGWEASTERLYPRDGVLGKGIHAKLKLDGSHYLVFVKSIYMAKK 185
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 PVKMPGYHYVDKRLDVTNNKDYTSVEQREISIAKRLPVACCFRVRKSRH 230
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 PVQLPGYHYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221

RESULT 14
US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; APPLICANT: Lukanov, Sergey
; APPLICANT: Yanushovich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Query Match 61.5%; Score 780; DB 14; Length 225;
Best Local Similarity 62.6%; Pred. No. 4.4e-76;
Matches 144; Conservative 28; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAQMTYKYVMGTVNGHYFEVGDGKGKPYEGEQTURLAVTTKGGPLPFAWDILSPQC 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 6 NVIKFMRPKVRMEGTVNGHFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSPQF 65
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 QXGSIPTFKYPEDIPDYVKQSPFGRYTWERIMNFEDGAVCTVSNDSIIQNCFCFIYHKFS 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 66 QXGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFIYKVKFI 125
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 GLNFPNGPVMOCKTQGWEPNTERLFARDGMLIGNFMALKLEGGHYLCBFKSTYKARK 180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 GVNFDGPFVMOCKTMGWEASTERLYPRDGVLGKGIHAKLKLDGSHYLVFVKSIYMAKK 185
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 PVKMPGYHYVDKRLDVTNNKDYTSVEQREISIAKRLPVACCFRVRKSRH 230
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 PVQLPGYHYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221

RESULT 13
US-10-006-922-12
; Sequence 12, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey A.
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
```

```
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-12

Query Match 61.5%; Score 780; DB 13; Length 225;
Best Local Similarity 62.6%; Pred. No. 4.4e-76;
Matches 144; Conservative 28; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAQMTYKYVMGTVNGHYFEVGDGKGKPYEGEQTURLAVTTKGGPLPFAWDILSPQC 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 6 NVIKFMRPKVRMEGTVNGHFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSPQF 65
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 QXGSIPTFKYPEDIPDYVKQSPFGRYTWERIMNFEDGAVCTVSNDSIIQNCFCFIYHKFS 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 66 QXGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFIYKVKFI 125
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 GLNFPNGPVMOCKTQGWEPNTERLFARDGMLIGNFMALKLEGGHYLCBFKSTYKARK 180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 GVNFDGPFVMOCKTMGWEASTERLYPRDGVLGKGIHAKLKLDGSHYLVFVKSIYMAKK 185
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 PVKMPGYHYVDKRLDVTNNKDYTSVEQREISIAKRLPVACCFRVRKSRH 230
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 PVQLPGYHYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221

RESULT 14
US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; APPLICANT: Lukanov, Sergey
; APPLICANT: Yanushovich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Query Match 61.5%; Score 780; DB 14; Length 225;
Best Local Similarity 62.6%; Pred. No. 4.4e-76;
Matches 144; Conservative 28; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAQMTYKYVMGTVNGHYFEVGDGKGKPYEGEQTURLAVTTKGGPLPFAWDILSPQC 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 6 NVIKFMRPKVRMEGTVNGHFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSPQF 65
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 QXGSIPTFKYPEDIPDYVKQSPFGRYTWERIMNFEDGAVCTVSNDSIIQNCFCFIYHKFS 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 66 QXGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFIYKVKFI 125
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 GLNFPNGPVMOCKTQGWEPNTERLFARDGMLIGNFMALKLEGGHYLCBFKSTYKARK 180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 GVNFDGPFVMOCKTMGWEASTERLYPRDGVLGKGIHAKLKLDGSHYLVFVKSIYMAKK 185
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 PVKMPGYHYVDKRLDVTNNKDYTSVEQREISIAKRLPVACCFRVRKSRH 230
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 PVQLPGYHYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221
```

```
Db 66 QYGSKVYVKKHADIPTYKLSFPEGKWERVNMNFDGGVTVTQDSSLQDGCIFYVKRFI 125
Qy 121 GINFPNGPVWQKKTQGWEPNTERLFARDGMLIGNNFWALKLEGGHYLCCEPKSTYKARK 180
Db 126 GVNFPDGFVWQKTMGWEASTERLYPRDGVLGKGIHAKLKLKDGGHYLVFEFKSIYMAKK 185
Qy 181 PVKMPGYHYVDRKLDVTNNHNDYTSVEQREISIAKKPLVACCFRVRKSRH 230
Db 186 PVQLPGYVYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221
```

```
RESULT 15
US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; FILE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CPI
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(225)
; OTHER INFORMATION: wild-type DeRed
US-10-121-258-1
```

```
Query Match 61.5%; Score 780; DB 14; Length 225;
Best Local Similarity 62.6%; Pred. No. 4.4e-76;
Matches 144; Conservative 28; Mismatches 44; Indels 14; Gaps 1;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVGDGKGPYEGEQTVRLAVTKGGPLPFANDILSPQC 60
Db 6 NVIKEFMEFKVMECTVNGHEFEIEGEGEPYEGHNTVKLVKYGGLPFANDILSPQF 65
Qy 61 QYGSIPFTKYBEDIPDYVQSFPGRYTWERIMNFDGAVCTVSDSSIQGNCFIYHVKPS 120
Db 66 QYGSKVYVKKHADIPTYKLSFPEGKWERVNMNFDGGVTVTQDSSLQDGCIFYVKRFI 125
Qy 121 GINFPNGPVWQKKTQGWEPNTERLFARDGMLIGNNFWALKLEGGHYLCCEPKSTYKARK 180
Db 126 GVNFPDGFVWQKTMGWEASTERLYPRDGVLGKGIHAKLKLKDGGHYLVFEFKSIYMAKK 185
Qy 181 PVKMPGYHYVDRKLDVTNNHNDYTSVEQREISIAKKPLVACCFRVRKSRH 230
Db 186 PVQLPGYVYVDSKLDITSHNEDYTIIVEQYE-----RTEGRH 221
```

Search completed: August 12, 2004, 06:51:21
Job time : 487.141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 24.1414 Seconds
(without alignments)
493.990 Million cell updates/sec

Title: US-09-890-463-3

Perfect score: 1268

Sequence: 1 SVIAQMTYKYVMSGTVNGH.....STARPLVACCFRKYSRHK 231

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/6C COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	482.5	38.1	238	3	US-09-277-716-16
2	482.5	38.1	238	4	US-09-609-161B-16
3	482.5	38.1	238	4	US-09-626-581D-65
4	482.5	38.1	238	4	US-09-415-763B-65
5	482.5	38.1	238	4	US-09-626-580C-65
6	473.5	37.3	238	3	US-09-277-716-32
7	473.5	37.3	238	4	US-09-609-161B-32
8	471.5	37.2	238	4	US-09-839-650-3
9	197.5	15.6	238	4	US-09-472-065A-5
10	192.5	15.2	238	4	US-09-479-645A-12
11	191.5	15.1	238	4	US-09-023-946B-28
12	189.5	14.9	238	4	US-09-023-946B-23
13	188.5	14.9	238	4	US-09-023-946B-36
14	188.5	14.9	240	4	US-09-129-192C-49
15	188	14.8	238	4	US-09-023-946B-22
16	187.5	14.8	238	1	US-08-337-915A-2
17	187.5	14.8	238	1	US-08-753-143-2
18	187.5	14.8	238	2	US-08-679-865-2
19	187.5	14.8	238	2	US-08-680-876-2
20	187.5	14.8	238	2	US-08-792-553-2
21	187.5	14.8	238	3	US-08-893-327-16
22	187.5	14.8	238	3	US-08-753-144-2
23	187.5	14.8	238	3	US-09-094-359-2
24	187.5	14.8	238	3	US-09-172-063-2
25	187.5	14.8	238	3	US-09-121-539-1
26	187.5	14.8	238	3	US-09-263-975-2
27	187.5	14.8	238	4	US-08-727-452-2

28	187.5	14.8	238	4	US-09-418-785-1	Sequence 1, Appli
29	187.5	14.8	238	4	US-09-214-909-2	Sequence 2, Appli
30	187.5	14.8	238	4	US-09-479-645A-10	Sequence 10, Appl
31	187.5	14.8	238	4	US-09-479-645A-159	Sequence 159, App
32	187.5	14.8	238	4	US-09-129-192C-2	Sequence 2, Appli
33	187.5	14.8	238	4	US-09-129-192C-74	Sequence 74, Appl
34	187.5	14.8	238	4	US-09-602-641-2	Sequence 2, Appli
35	187.5	14.8	238	4	US-09-704-463-2	Sequence 2, Appli
36	187.5	14.8	238	4	US-09-472-065A-4	Sequence 4, Appli
37	187.5	14.8	238	4	US-09-472-065A-6	Sequence 6, Appli
38	187.5	14.8	238	4	US-09-023-946B-29	Sequence 29, Appl
39	187.5	14.8	238	4	US-09-920-922-4	Sequence 4, Appli
40	187.5	14.8	238	5	PCT-US95-14692-2	Sequence 2, Appli
41	187.5	14.8	239	3	US-08-646-538-2	Sequence 2, Appli
42	187.5	14.8	239	3	US-09-094-359-4	Sequence 4, Appli
43	187.5	14.8	239	3	US-09-172-063-3	Sequence 3, Appli
44	187.5	14.8	239	3	US-09-503-223-2	Sequence 2, Appli
45	187.5	14.8	239	4	US-09-513-783A-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1

US-09-277-716-16

; Sequence 16, Application US/09277716A

; Patent No. 6232107

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

; CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: US/09/277,716A

; EARLIER FILING DATE: 1998-10-01

; EARLIER APPLICATION NUMBER: 60/089,367

; EARLIER FILING DATE: 1998-06-15

; EARLIER APPLICATION NUMBER: 60/079,624

; EARLIER FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 16

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Renilla mulleri

; FEATURE:

; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Query Match 38.1%; Score 482.5; DB 3; Length 238;

Best Local Similarity 44.0%; Pred. No. 2.2e-45;

Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

QY 7 MTKYVMSGTVNGHYPEVGDGKPKYEGEQVRLAVTKGGLPFPFADWILSPQCOYGSIP 66

Db 15 MSKYVNLGIVNNHVTFTMEGCKGNILFGNLQVRLVTKGALPFPFADIVSFAFYGNRT 74

QY 67 FTKYPEDIPDYKQSPFGRYTWTWRIMNFDGAVCTVSDSSIQGNCFIYHVKFSGLNFPFP 126

Db 75 FTKYPNDISDYFIQSPFAGFVETLRYEDGGLVEIRSDINLIEDKFVYRVEYKSGNFPD 134

QY 127 NSGVWOKKTQGWEPNTERLFARDGMLIGNNFALLKGGHYLCBFKSYKARKPKV-MP 185

Db 135 DGFVWQKTLTGIEPSPFAMVMNNGVLGVILVYKLSGKYSCHMKTLMLKSGVVKFP 194

QY 186 GYHYVDRLDVTNHNKDYTSVEQRETSIAR-----KPL 218

Db 195 SYHFIQRLKLT-YVEDGGFVEQHEHTAQMTSICKPL 231

RESULT 2

US-09-609-161B-16

```

; Sequence 16, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-16

Query Match      38.1%; Score 482.5; DB 4; Length 238;
Best Local Similarity 44.0%; Pred. No. 2.2e-45;
Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

QY 7 MTKYKMGSTVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQCQYGSIP 66
Db 15 MSYKVNLEGIVNNHVFTEGCGKGNILFGNQLVQIRVTKGAPLPAFAFADIVSPAFOYGNRT 74

QY 67 FTKYPEDIPDYVKQSPGRTYRIMNPFEDGAVCTVSDSSIQGNCFIYHVKFSGLNPP 126
Db 75 FTKYENDISDFIQSPFAGFYRILRYEDGGLVIRSDINLIEDKFVYRVYKSGSNFPD 134

QY 127 NGPVNMQKKTQGWEPNTERLFARDGMLIGNNFMALKEGGHYLCBFKSTYKARKPKV-MP 185
Db 135 DGPVMQKTLIGLEPSPFEAMYNMNGVLGVEVILVYKLSNGKYSCHMKTLKMSKGVVKEFP 194

QY 186 GHYVDRKLDVTNNHNDYTSVEQREISAR-----KPL 218
Db 195 SYHFIQHRLEKT-YVEDGGFVEQHETAIAQMTSIGKPL 231

RESULT 3
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-09-626-581D-65

Query Match      38.1%; Score 482.5; DB 4; Length 238;
Best Local Similarity 44.0%; Pred. No. 2.2e-45;
Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

QY 7 MTKYKMGSTVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQCQYGSIP 66
Db 15 MSYKVNLEGIVNNHVFTEGCGKGNILFGNQLVQIRVTKGAPLPAFAFADIVSPAFOYGNRT 74

QY 67 FTKYPEDIPDYVKQSPGRTYRIMNPFEDGAVCTVSDSSIQGNCFIYHVKFSGLNPP 126
Db 75 FTKYENDISDFIQSPFAGFYRILRYEDGGLVIRSDINLIEDKFVYRVYKSGSNFPD 134

QY 127 NGPVNMQKKTQGWEPNTERLFARDGMLIGNNFMALKEGGHYLCBFKSTYKARKPKV-MP 185
Db 135 DGPVMQKTLIGLEPSPFEAMYNMNGVLGVEVILVYKLSNGKYSCHMKTLKMSKGVVKEFP 194

QY 186 GHYVDRKLDVTNNHNDYTSVEQREISAR-----KPL 218
Db 195 SYHFIQHRLEKT-YVEDGGFVEQHETAIAQMTSIGKPL 231

RESULT 4
US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-09-415-765B-65

Query Match      38.1%; Score 482.5; DB 4; Length 238;
Best Local Similarity 44.0%; Pred. No. 2.2e-45;
Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

QY 7 MTKYKMGSTVNGHYFEVGGDKGPKYEGEQTIVRLAVTKGGPLPFAWDILSPQCQYGSIP 66
Db 15 MSYKVNLEGIVNNHVFTEGCGKGNILFGNQLVQIRVTKGAPLPAFAFADIVSPAFOYGNRT 74

QY 67 FTKYPEDIPDYVKQSPGRTYRIMNPFEDGAVCTVSDSSIQGNCFIYHVKFSGLNPP 126
Db 75 FTKYENDISDFIQSPFAGFYRILRYEDGGLVIRSDINLIEDKFVYRVYKSGSNFPD 134

QY 127 NGPVNMQKKTQGWEPNTERLFARDGMLIGNNFMALKEGGHYLCBFKSTYKARKPKV-MP 185
Db 135 DGPVMQKTLIGLEPSPFEAMYNMNGVLGVEVILVYKLSNGKYSCHMKTLKMSKGVVKEFP 194

QY 186 GHYVDRKLDVTNNHNDYTSVEQREISAR-----KPL 218
Db 195 SYHFIQHRLEKT-YVEDGGFVEQHETAIAQMTSIGKPL 231

RESULT 5
US-09-626-580C-65
; Sequence 65, Application US/09626580C
; Patent No. 6562617
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES

```

```

RESULT 8
US-09-839-650-3
; Sequence 3, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Strategene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein

```

```
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-839-650-3

Query Match      37.2%; Score 471.5; DB 4; Length 238;
Best Local Similarity 43.6%; Pred. No. 3.7e-44;
Matches 95; Conservative 40; Mismatches 76; Indels 7; Gaps 3;

QY 7 MTYKVMSTGVNGHYFEVGGKPYEGEQTIVRLAVTKGGPLPPAWDTLSPQCQXGSIPTKY 66
Db 15 MSYKVNLEGI VNNHVTVEGGRKNILFGNQLVHRTVTKGGPLPPAFDIVSAFYQGNRT 74
QY 67 FTKYPEDIPYVKQSPGRTYTWERIMNFEDGAVCTVSDSSIQGNCFIYHVKFSGLNFP 126
Db 75 FTKYPNDISDFIQSFPAGFMVYERTLRYEDGGLVEIRSDINLIEDKFVIRVEYKGSNFPD 134
QY 127 NGPVMOKKTQGWEPNTERLFPADGMLIGNFMALKLEGGHYLCFCKSTYKARKPKV-MP 185
Db 135 DGPVMQKTLTGIEPSEAFMYMNNGVLVGEVILVYKLSNGKYSCHMKTLMKSKGVVKEFP 194
QY 186 GYHYVDRLDVNHNKDYTSVEQREISAR-----KPL 218
Db 195 SYHFQHRLEKT-YVEDGQFVEQHEHTAQTASIGKPL 231

RESULT 9
US-09-472-065A-5
; Sequence 5, Application US/09472065A
; Patent No. 6638732
; GENERAL INFORMATION:
; APPLICANT: Evans, Krista
; TITLE OF INVENTION: Mutants of Green Fluorescent Protein
; FILE REFERENCE: 0942.4020002
; CURRENT APPLICATION NUMBER: US/09/472,065A
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/970,762
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/030,935
; PRIOR FILING DATE: 1996-11-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria, A1 mutant
US-09-472-065A-5

Query Match      15.6%; Score 197.5; DB 4; Length 238;
Best Local Similarity 25.7%; Pred. No. 9.5e-14;
Matches 53; Conservative 46; Mismatches 90; Indels 17; Gaps 8;

QY 11 VYMSGTVNGHYFEVGGKPYEGEQTIVRLAVTKGGPLPPAWDTLSPQCQXGSIPTKY 70
Db 16 VELGDVNGHKFVS GEGEGDATYGLTKFICIT-GKLPVWPVLVTTCAYGVQCFSKY 74
QY 71 PEDIP--DYVKQSPGRTYTWERIMNFEDGAVCTVSDSSIQGNCFIYHVKFSGLNFP 128
Db 75 PDHMKQHDFFKSA MPEGVVQERTIFFKDDGNKYKTRAEVKFEGDTLVNRIELKGIDFKEDG 134
QY 129 PVMQKKTQGWEPNTERLFP-----ARDGMLIGNFMALKLEGGHYLCF--KSTYKARK 180
Db 135 NILGHKLE-YNYSNHNVMADKQNGIKV--NFKIRHNIEDGVSQVLADHYQOQNTPIGDG 191
QY 181 PVKMPGYHYVDRLKLDVT---NHNKDY 203
Db 192 PVLLPDNHYLSQSALS KDPNEKRDH 217

RESULT 11
US-09-023-946B-28
; Sequence 28, Application US/09023946B
; Patent No. 6670449
; GENERAL INFORMATION:
; APPLICANT: GERO MIESENBOCK, ET AL.
; TITLE OF INVENTION: HYBRID MOLECULES AND THEIR USE FOR OPTICALLY DETECTING CHANGES IN CELLULAR MICROENVIRONMENTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
```


Db 192 FVLLPDNHYL 201

Search completed: August 12, 2004, 06:21:06
Job time : 25.1414 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 21.6701 Seconds
(without alignments)
1043.144 Million cell updates/sec

Title: US-09-890-463-4

Perfect score: 1287

Sequence: 1 SVIAKQWYKYVMSTGVNH.....KPVVACRFVRKSRHKYAVA 235

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	185.5	14.4	238	1 JQ1514	green-fluorescent
2	97	7.5	583	2 S64909	probable membrane
3	86.5	6.7	368	2 A81289	hypothetical prote
4	86	6.7	770	2 S00643	anthranilate synth
5	84	6.5	725	1 IUMSNG	neural cell adhes
6	84	6.5	787	2 E82323	organic solvent to
7	83	6.4	260	2 B70419	hypothetical prote
8	83	6.4	752	2 S23818	hypothetical prote
9	82.5	6.4	600	2 T34757	probable oligopept
10	80.5	6.3	390	2 D64903	probable arylsulfa
11	80.5	6.3	3461	2 S58870	reelin precursor -
12	79.5	6.2	876	2 A89944	alaninyl-TRNA synth
13	79	6.1	403	2 T51828	probable photosyst
14	79	6.1	687	1 B32382	ubiquinol-cytochr
15	79	6.1	898	1 DUBPT4	DNA-directed DNA p
16	79	6.1	5175	2 T20992	hypothetical prote
17	79	6.1	5198	2 T43290	hemocytin precurs
18	78.5	6.1	363	2 S30149	cysteine proteinas
19	78.5	6.1	390	2 F90891	probable enzyme li
20	78.5	6.1	390	2 C85726	probable enzyme 22
21	78.5	6.1	870	2 T47454	lipoygenase AtLOX
22	78.5	6.1	896	2 JQ2391	lipoygenase (EC 1
23	78	6.1	334	2 D95982	hypothetical prote
24	78	6.1	380	2 T09683	hypothetical prote
25	78	6.1	642	2 T29757	protein UNC-89 - C
26	77.5	6.0	207	2 A48608	E1 glycoprotein -
27	77.5	6.0	266	1 CDFU13	chlorophyll a/b-bi
28	77.5	6.0	297	2 E71425	hypothetical prote
29	77.5	6.0	333	2 B64380	hypothetical prote

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C>Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001

C:Accession: J50692; JQ1514; FQ0335; S48693; S51330; S51331

R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cornier, M.J.

Gene 111, 229-233, 1992

A>Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: JQ1514; MUID:92175527; PMID:1347277

A:Accession: J50692

A:Molecule type: DNA

A:Residues: 1-107, 'S', 109-238 <PRA1>

A:Cross-references: GB:M62654; MID:g155662; PIDN:AAA27722.1; PID:g155663

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99, 'P', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>

A:Cross-references: GB:M62653; MID:g155660; PIDN:AAA27721.1; PID:g155661

A:Accession: FQ0335

A:Molecule type: protein

A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A>Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>

A:Cross-references: GB:L29345; MID:g606383; PIDN:AAA58246.1; PID:g606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13, 'V', 15-24, 'O', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R', 2

A:Cross-references: EMBL:X83959; MID:g634008; PIDN:CAAS8789.1; PID:g634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q', 2

A:Cross-references: EMBL:X83960; MID:g634010; PIDN:CAAS8790.1; PID:g634011

A:Experimental source: Clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GLF

A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-95

A>Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A>Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

probable TonB-depe
genome polyprotein
structural polypro
protein kinase, pr
collagen alpha 2 c
xylokinase homol
68K protein - phag
WD-40 repeat prote
chlorophyll a/b-bi
repa protein - Bsc
endo-1,4-beta-xyla
hypothetical prote
diptide ABC tran
RNA-directed RNA p
E2 glycoprotein pr
alpha-1-B-glycopro

30 77.5 6.0 690 2 AB0124
31 77.5 6.0 1241 2 S26373
32 77.5 6.0 1242 2 S72350
33 77 6.0 554 2 F90399
34 77 6.0 1822 2 S63985
35 76.5 5.9 487 2 E69895
36 76.5 5.9 613 2 S19676
37 76.5 5.9 1526 2 AC2239
38 76 5.9 266 1 CDFJ25
39 76 5.9 288 2 S5184
40 76 5.9 342 2 A60154
41 76 5.9 444 2 E65045
42 76 5.9 658 2 H84385
43 76 5.9 725 1 P21VBS
44 76 5.9 1162 2 S14933
45 75.5 5.9 237 2 A42013

C;Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane

F;1-19/Domain: signal sequence #status predicted <SIG>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN3A>
F;635-685/Domain: fibronectin type III repeat homology <FN3B>
F;41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted
F;222,316,348,424,450,479/Binding site: carbohydrate (Asn) #status predicted

Query Match 6.5%; Score 84; DB 1; Length 725;
Best Local Similarity 22.9%; Pred. No. 20;
Matches 54; Conservative 38; Mismatches 82; Indels 62; Gaps 16;

Qy 10 KVMGSGTVNGHYFVBEVGGK--PYE-----GEQTVR-LAVTKGGPLPFAWDILSP- 58
Db 455 KII--NTPSASYLEVTDPSENDFGNYNTAVNRIGQESLEFLVQADTSPSSIDRVEPV 512
Qy 59 ----QCQY-----GSTPFTKVPEDIPYVQKSPFGRYTWERIMNPDGAVCTVSDSSI 108
Db 513 SSTAQVQFDEPATGGVPILKYKAEWKSGLGESW--HFTWYD-----AKEANN 558
Qy 109 QGNCFI-----YHVKFGSLNPPNPVMQ---KKTQG---WEPNTERLFARDGMLI 153
Db 559 EGIVTIMGLKPEITYSDRLAALNGKLGELIQWPSSEKTPQVPPELSAPKLEGOMGEDGNSI 618
Qy 154 GNNFMALKLEGG---HYLCEPKSTYKAKXP-KRMP-GYHYVDRKLDVTNHNKDY 203
Db 619 KVN--IKQDDGGSPIRHYLVKYRALASEWKPEIRLPSGSHV--MLKSLDWNAAEY 670

RESULT 6
E82323
organic solvent tolerance protein VC0446 [imported] - Vibrio cholerae (strain N16961 sero
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: E82323
F;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: E82323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-787 <HEI>
A;Cross-references: GB:AE004131; GB:AE003852; NID:g9654856; PIDN:AAF93619.1; GSPDB:GN0012
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0446
A;Map position: 1
C;Superfamily: organic solvent tolerance protein

Query Match 6.5%; Score 84; DB 2; Length 787;
Best Local Similarity 18.8%; Pred. No. 22;
Matches 44; Conservative 34; Mismatches 92; Indels 64; Gaps 9;

Qy 6 QMTYKVMYSGTVN-----GHYEVEGDGKGKPYEGEQTVPLAVTKGGPLPFA--WDIL 56
Db 388 QLSYNYAPEPMKYMLDLDLVSHVSRRFTDARKP-----SATRVHTEPGLKIPFSNTWGNW 443
Qy 57 SPOCQYGSIPPTKYPEDIPYVQKSPFGRYTWERIMNPDGAVCTVSDSSISQGNCFIYH 116
Db 444 TTEAR---VLGTYQQDLKTTD-----AKLEESVTRVPIERSV----- 480
Qy 117 VKFSLGNFPNPPGPMYKQKTQGWEPNTERLFA-----RDGMLIGNNFMAL----- 160

Db 461 DKHQEGIIPEKIAKVFFQRKDWKKMFEEE-----MNAEAIKKIIMKGAGS--CSTKPEV 513

QY 177 KAKKPVKMPGYHYVDRKLDVTNHNKDYTS-----VEQCE 210

Db 514 E-----RYVKFSDDLNLNLSNYTESVLSLIEECE 543

514 E-----RYVKFSDDFLNELSNYTEVLSNLSIECE 543

Search completed: August 12, 2004, 06:14:03
Job time : 25.6701 secs

Job time : 25.6701 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 14.4467 Seconds
(without alignments)
847.008 Million cell updates/sec

Title: US-09-890-463-4
Perfect score: 1287

Sequence: 1 SVIAKQMTYKYVMSGTVNGH.....KPVVACRFVRKSRKYAVA 235

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	188.5	14.6	238	1	GFP_AEQVI
2	86	6.7	770	1	TRPG_ASPNG
3	84	6.5	725	1	NCA2_MOUSE
4	84	6.5	787	1	OSTA_VIBCH
5	83	6.4	260	1	Y069_AQUAE
6	81	6.3	1142	1	ENAM_HUMAN
7	80.5	6.3	385	1	YDEM_ECOLI
8	80.5	6.3	3461	1	REIN_MOUSE
9	79.5	6.2	876	1	SYA_STAAM
10	79.5	6.2	876	1	SYA_STAAM
11	79	6.1	403	1	H136_ARATH
12	79	6.1	687	1	CYBC_BRAVA
13	79	6.1	898	1	DPOI_BPT4
14	78.5	6.1	896	1	LOXC_ARATH
15	78.5	6.1	3462	1	REIN_RAT
16	78	6.1	6632	1	UN89_CAEEL
17	77.5	6.0	266	1	CB21_PETSP
18	77.5	6.0	333	1	Y642_METUA
19	77.5	6.0	366	1	MUTA_BUCAP
20	77	6.0	726	1	RRP2_INBP9
21	76.5	5.9	1526	1	YV46_ANASP
22	76	5.9	266	1	CB24_PETSP
23	76	5.9	342	1	XVNA_CALSA
24	76	5.9	444	1	YGNF_ECOLI
25	76	5.9	725	1	RRP2_INBSI
26	76	5.9	1162	1	VGL2_IBVB
27	75.5	5.9	587	1	COAT_PAVL3
28	75	5.8	248	1	YA14_SCHPO
29	75	5.8	267	1	CB22_PETSP
30	75	5.8	392	1	ANIA_NEIGO
31	75	5.8	905	1	HXA1_HAEIN
32	75	5.8	1004	1	MV10_MOUSE
33	74.5	5.8	513	1	YHCX_BACSU

34	74.5	5.8	2390	1	SPCP_HUMAN	O15020 homo sapien
35	74	5.7	265	1	CB2B_LYCES	P07370 lycopersico
36	74	5.7	267	1	CB25_PETSP	P04783 petunia sp.
37	74	5.7	277	1	YF99_METUA	Q58994 methanococ
38	74	5.7	398	1	Y4QJ_RHISN	P55631 rhizobium s
39	74	5.7	800	1	P727_YEAST	P32606 saccharomyc
40	73.5	5.7	266	1	CB21_SINAL	P13851 sinapis alb
41	73.5	5.7	578	1	NRD2_BAT	Q63504 rattus norv
42	73.5	5.7	640	1	MYB_BOVIN	P46200 bos taurus
43	73.5	5.7	1142	1	ENAM_PIG	O97939 sus scrofa
44	73.5	5.7	1254	1	UBPC_YEAST	P39538 saccharomyc
45	73	5.7	213	1	GUN5_HUMIN	P43316 humicola in

ALIGNMENTS

RESULT 1
GFP_AEQVI
ID GFP_AEQVI STANDARD; PRT; 238 AA.
AC P42212; Q17104; Q27903;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Green fluorescent protein.
GN GFP
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92175527; PubMed=1347277;
RA Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,
RA Cormier M.J.;
RT "Primary structure of the Aequorea victoria green-fluorescent
RT protein.";
RL Gene 111:229-233 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94185810; PubMed=8137953;
RA Inouye S., Tsuji F.I.;
RT "Aequorea green fluorescent protein. Expression of the gene and
RT fluorescence characteristics of the recombinant protein.";
RL FEBS Lett. 341:277-280 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rowland G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage.";
RL Plant Mol. Biol. 33:989-999 (1997).
RN [4]
RP CHROMOPHORE.
RX MEDLINE=93192221; PubMed=8448132;
RA Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;
RT "Chemical structure of the hexapeptide chromophore of the Aequorea
RT green-fluorescent protein.";
RL Biochemistry 32:1212-1218 (1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96355665; PubMed=8703075;
RA Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,
RA Remington S.J.;
RT "Crystal structure of the Aequorea victoria green fluorescent
RT protein.";
RL Science 273:1392-1395 (1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98294543; PubMed=9631087;
RA Yang F., Moss L.G., Phillips G.N. Jr.;
RT "The molecular structure of green fluorescent protein.";
RL Nat. Biotechnol. 14:1246-1251 (1996).

192 PVLPLPDHNYLSTQSALSQDPNEKRHD 217

```
DR PRINTS; PR00099; CPSEATASE.
DR PRINTS; PR00096; GATASE.
DR PRODOM; PD001511; IGFS; 1.
DR TIGRFAMS; TIGR00566; trpg_papA; 1.
DR PROSITE; PS00442; GATASE_TYPE_1; 1.
DR PROSITE; PS00614; IGFS; 1.
KW Tryptophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
KW Decarboxylase; Transferase; Glutamine amidotransferase.
FT DOMAIN 25 219 GLUTAMINE AMIDOTRANSFERASE.
FT DOMAIN 255 519 INDOL-3-GLYCEROL PHOSPHATE SYNTHASE.
FT ACT_SITE 104 104 GATASE (BY SIMILARITY).
FT ACT_SITE 199 199 GATASE (BY SIMILARITY).
FT ACT_SITE 201 201 GATASE (BY SIMILARITY).
SQ SEQUENCE 770 AA; 82909 MW; 3608DE5B23097012 CRC64;

Query Match 6.7%; Score 86; DB 1; Length 770;
Best Local Similarity 24.3%; Pred. No. 5.4;
Matches 50; Conservative 23; Mismatches 61; Indels 72; Gaps 12;

QY 10 KYMSGTV-NGHYFEVGDGKGYEGEQTVRLAVTKGGLPFPADILSPQCVGSIPT 68
DB 117 KDVVTGEILHGKTSPLKHDGKG-AYEG-----LP-----GSLAVT 150
QY 69 KYPE-----DIPYVKOSPPGRYTWERIMNFEDGAVCTVSDSSIQGNCFIYHVKFS-- 120
DB 151 RYHSLAGTHATIPDCLEVS-----SSVQADD-----SNKDVIMG---VRHKKLAVE 194
QY 121 GLNFPFNGPVNMQKKTQGWEPNTERLFDGDMGLIGNNFMAKL---EGGCHYLCEFKSTYK 177
DB 195 GVQFHP-----ESILTEYGRIMFERNFLKLTAGTWEGNGKHFGESSTTK 238
QY 178 AKKPVKMPGYHYVDKLDVTHNKDY 203
DB 239 ATVPSNPP--PKTDKLSILERYIDH 262

RESULT 3
NCAR2 MOUSE
ID NCAR2_MOUSE STANDARD; PRT; 725 AA.
AC P13594; Q61950;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE (NCAM-120).
GN NCAM1 OR NCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=87246524; PubMed=3595563;
RA Barthelemy D., Santoni M.-J., Wille W., Ruppert C., Caix J.-C.,
RA Hirsch W.-R., Fontecilla-Camps J.-C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
a Mr 79,000 polypeptide without a membrane-spanning region.";
RL EMBO J. 6:907-914 (1987).
RN [2]
RP SEQUENCE OF 20-700 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=89251563; PubMed=2721486;
RA Santoni M.J., Barthelemy D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
generates at least eight types of NCAM cDNA in mouse brain.";
RL EMBO J. 8:385-392 (1989).
RN [3]
RP SEQUENCE OF 642-725 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=88283628; PubMed=3396534;
RA Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
```


R		GROWTH; GROWTH; GROWTH.
R	MIM; 606585; -	C:extracellular matrix; NAS.
R	GO; GO:0005575;	GO; GO:0003045; F:structural constituent of tooth enamel; NAS.
R	GO; GO:0030345;	P:bone mineralization; NAS.
R	GO; GO:0030282;	P:odontogenesis; NAS.
R	GO; GO:0042476;	Biom Mineralization; Extracellular matrix; Glycoprotein; Signal.
W	SIGNAL	1 39 POTENTIAL.
T	CHAIN	40 1142 ENAMELIN.
T	CARBOHYD	114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
T	CARBOHYD	126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
T	CARBOHYD	245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
T	CARBOHYD	252 252 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 934 934 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1142 AA; 128745 MW; 77419C4375EAD6EC CRC64;
 Query Match 6.3%; Score 81; DB 1; Length 1142;
 Best Local Similarity 29.5%; Pred. No. 24;
 Matches 23; Conservative 12; Mismatches 35; Indels 8; Gaps 3;
 Qy 18 NGHVEVEGDKGKPYEGEOTVRLVATYKG-----GPIPFWDILSPQCQYGSIBFTKYPE 72
 Db 493 NSYY--PRGDSRKVPNSDGTQSQNLPGKIVLGSRMPYSESTNQSELKHSSYPAPVPE 550
 Qy 73 DIPDYVKQSEP-GRYTWE 89
 Db 551 EIPSPAKEHFFAGRNWTD 568
 RESULT 7
 YDEM_ECOLI
 ID_YDEM_ECOLI STANDARD; PRT; 385 AA.
 AC P76134; E77755;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ydem.
 GN YDEM OR B1497.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasundaram S.,
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC -!- SIMILARITY: BELONGS TO THE ASL6/ATSB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000247; AAC74570.1; ALT_INIT.
 DR EMBL; D90791; BAA15168.1; -.
 DR EMBL; D90792; BAA15171.1; -.
 DR EcoGene; EG13795; ydem.
 DR InterPro; IPR007197; Radical_SAM.
 DR Pfam; PF04055; Radical_SAM; I.

KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 385 AA; 44518 MW; 964E34F73E680329 CRC64;
 Query Match 6.3%; Score 80.5; DB 1; Length 385;
 Best Local Similarity 19.4%; Pred. No. 7.4; 103; Indels 75; Gaps 12;
 Matches 54; Conservative 46; Mismatches 46; Mismatches 46; Mismatches 46;
 Qy 10 KVMSTGTVNGHYFEVGDGKGP-----YEGQTVRLAV-TKGGPLPF 51
 Db 44 KQYIAASGNQVYFTWQ--GEPTLAGLDPRKVIHYQORVAGQKRFNALQTNGLNN 100
 Qy 52 AWDILSPQCQYGSIBFTKYPEIDPYVKQSPGRTW-----ERIMNFE-DGAVCTVS 103
 Db 101 EWCAFLKEHEFLVGISIDGQELHVDYRRNSNGTFAKVIAAIAERLKSQVEFNLTVI 160
 Qy 104 NDSSIQNCFTYH-----VKFSG-----LNFPN--GP 129
 Db 161 NNVNHYPLEVIYHFLKSGKMQFIELLETGTNTIDFSHSENTFRIDSVPTAYCK 220
 Qy 130 VMQKKTQWEPN-TERLFARDGMLIGNFMALKEGGHYLCEPKSTYKAKKPKVMPG-- 186
 Db 221 FMSTIFMQWVKNDVGEIFIRQFESFVSRL-----GNHTSCIFQESCKDLVVESNGDI 275
 Qy 187 ---YHYVDKLDVTNHNK-DYTSVEQCEISARIPVVA 220
 Db 276 YECDHFVYPOYKIGNINKSELKTNVSLTAQKKRIPA 313
 RESULT 8
 RELN_MOUSE
 ID_RELN_MOUSE STANDARD; PRT; 3461 AA.
 AC Q60841; Q9CUA6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Reelin precursor (EC 3.4.21.-) (Reeler protein).
 GN RELN OR RL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Cerebellum;
 RX MEDLINE=95231649; PubMed=7715726;
 RA D'Arcangelo G., Miao G.G., Chen S.-C., Soares H.D., Morgan J.I.,
 RA Curran T.;
 RT "A protein related to extracellular matrix proteins deleted in the
 RT mouse mutant reeler.";
 RL Nature 374:719-723(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=98086481; PubMed=9417911;
 RA Royaux I., Lambert de Rouvrol C., D'Arcangelo G., Demirov D.,
 RA Goffinet A.M.;
 RT "Genomic organization of the mouse reelin gene.";
 RL Genomics 46:240-250(1997).
 RN [3]
 RP SEQUENCE OF 2152-3461 FROM N.A. (ISOFORM 1).
 RC STRAIN=BALB/c; TISSUE=Brain;
 RX MEDLINE=95375789; PubMed=7647795;
 RA Hirotsune S., Takahara T., Sasaki N., Hirose K., Yoshiki A.,
 RA Ohashi T., Kusakabe M., Murakami Y., Muramatsu M., Watanabe S.,
 RA Nakao K., Katsuki M., Hayashizaki Y.;
 RT "The reeler gene encodes a protein with an EGF-like motif expressed by
 RT pioneer neurons.";
 RL Nat. Genet. 10:77-83(1995).
 RN [4]
 RP SEQUENCE OF 3044-3461 FROM N.A. (ISOFORM 2).
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97141547; PubMed=8987733;
RA D'Arcangelo G., Nakajima K., Miyata T., Ogawa M., Mikoshiba K.,
RA Curran T.,
RT "Reelin is a secreted glycoprotein recognized by the CR-50 monoclonal
RT antibody.";
RL J. Neurosci. 17:23-31(1997).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=21634904; PubMed=11689558;
RA Quattrocchi C.C., Wanneres F., Persico A.M., Ciafre S.A.,
RA D'Arcangelo G., Farace M.G., Keller P.,
RT "Reelin is a serine protease of the extracellular matrix.";
RL J. Biol. Chem. 277:303-309(2002).
RN [7]
RP TISSUE SPECIFICITY.
RX MEDLINE=97325946; PubMed=9182958;
RA Schiffmann S.N., Bernier B., Goffinet A.M.;
RT "Reelin mRNA expression during mouse brain development.";
RL Eur. J. Neurosci. 9:1055-1071(1997).
RN [8]
RP ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RX MEDLINE=99263436; PubMed=10328932;
RA Lambert de Rouvroit C., Bernier B., Royaux I., de Bergcyck V.,
RA Goffinet A.M.,
RT "Evolutionarily conserved, alternative splicing of reelin during brain
RT development.";
RL Exp. Neurol. 156:229-238(1999).
RN [9]
RP BINDING TO VLDLR AND ApoER2.
RX MEDLINE=20036019; PubMed=10571241;
RA Hiesberger T., Trommsdorff M., Howell B.W., Goffinet A.M., Mumby M.C.,
RA Cooper J.A., Herz J.,
RT "Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces
RT tyrosine phosphorylation of disabled-1 and modulates tau
RT phosphorylation.";
RL Neuron 24:481-489(1999).
RN [10]
RP FUNCTION.
RX MEDLINE=20359755; PubMed=10880573;
RA Yip J.W., Yip Y.P.L., Nakajima K., Capriotti C.;
RA "Reelin controls position of autonomic neurons in the spinal cord.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:8612-8616(2000).
CC -!- FUNCTION: Extracellular matrix serine protease that plays a role
CC in layering of neurons in the cerebral cortex and cerebellum.
CC Regulates microtubule function in neurons and neuronal migration.
CC Affects migration of sympathetic preganglionic neurons in the
CC spinal cord, where it seems to act as a barrier to neuronal
CC migration. Enzymatic activity is important for the modulation of
CC cell adhesion. Binding to the extracellular domains of lipoprotein
CC receptors VLDLR and ApoER2 induces tyrosine phosphorylation of
CC Dab1 and modulation of Tau phosphorylation.
CC -!- SUBUNIT: Binds to the ectodomains of VLDLR and ApoER2.
CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q60841-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q60841-2; Sequence=VSP_005577;
CC Name=3;
CC IsoId=Q60841-3; Sequence=VSP_005578;
CC -!- TISSUE SPECIFICITY: The major isoform 1 is neuron-specific. It is
CC abundantly produced during brain ontogenesis by the Cajal-Retzius
CC cells and other pioneer neurons located in the telencephalic
CC marginal zone and by granule cells of the external granular layer
CC of the cerebellum. Expression is located in deeper layers in the
CC developing hippocampus and olfactory bulb, low levels of
CC expression are also detected in the immature striatum. At early
CC developmental stages, expressed also in hypothalamic
CC differentiation fields, tectum and spinal cord. A moderate to low
CC level of expression occurs in the septal area, striatal fields,
CC habenular nuclei, some thalamic nuclei, particularly the lateral
CC geniculate, the retina and some nuclei of the reticular formation
CC in the central field of the medulla. Very low levels found in
CC liver and kidney. No expression in radial glial cells, cortical
CC plate, Purkinje cells and inferior olivary neurons. The minor
CC isoform 2 is only expressed in non neuronal cells. The minor
CC isoform 3 is found in the same cells as isoform 1, but is almost
CC undetectable in retina and brain stem.
CC -!- DEVELOPMENTAL STAGE: First detected at embryonic day 11.5.
CC Expression increases up to birth and remains high from post-natal
CC day 2 to 11 in both cerebellum and fore/midbrain. Expression
CC declines thereafter and is largely brain specific in the adult.
CC -!- DOMAIN: The basic C-terminal region is essential for secretion.
CC -!- PTM: N-glycosylated and to a lesser extent also O-glycosylated.
CC -!- DISEASE: Defects in reelin are the cause of the autosomal recessive
CC reeler (rl) phenotype which is characterized by impaired motor
CC coordination, tremors and ataxia. Neurons in affected mice fail to
CC reach their correct locations in the developing brain, disrupting
CC the organization of the cerebellar and cerebral cortices and other
CC laminated regions.
CC -!- SIMILARITY: Belongs to the reelin family.
CC -!- SIMILARITY: Contains 8 EGF-like domains.
CC -!- SIMILARITY: Contains 15 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U24703; AB91599.1; -;
CC EMBL; D63520; BAA09788.1; ALT_INIT.
CC EMBL; AK017094; BAB30592.1; -;
CC MGD; MGI:103022; ReIn.
CC GO; GO:0005615; C:extracellular space; IDA.
CC GO; GO:0007420; P:brain development; IMP.
CC GO; GO:0016477; P:cell migration; IMP.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR002860; GH_BNR.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR002861; Reeler.
CC Pfam; PF02012; BNR; 15.
CC Pfam; PF00008; EGF; 3.
CC Pfam; PF02014; Reeler; 1.
CC SMART; SM00181; EGF; 5.
CC PROSITE; PS00022; EGF_1; 7.
CC PROSITE; PS01186; EGF_2; 6.
CC PROSITE; PS50026; EGF_3; 5.
CC Hydrolase; Serine protease; Developmental protein; Matrix protein;
CC Cell adhesion; EGF-like domain; Glycoprotein; Repeat; Signal;
CC Alternative splicing.
CC SIGNAL 1 26 POTENTIAL.
CC CHAIN 27 3461 REELIN.
CC FT

FT DOMAIN 40 172 REBLER.
 FT DOMAIN 671 702 EGF-LIKE 1.
 FT DOMAIN 1030 1061 EGF-LIKE 2.
 FT DOMAIN 1409 1442 EGF-LIKE 3.
 FT DOMAIN 1765 1796 EGF-LIKE 4.
 FT DOMAIN 2129 2161 EGF-LIKE 5.
 FT DOMAIN 2478 2509 EGF-LIKE 6.
 FT DOMAIN 2853 2884 EGF-LIKE 7.
 FT DOMAIN 3228 3260 EGF-LIKE 8.
 FT REPEAT 593 604 BNR 1.
 FT REPEAT 799 810 BNR 2.

Query Match 6.3%; Score 80.5; DB 1; Length 3461;
 Best Local Similarity 19.9%; Pred. No. 99;
 Matches 56; Conservative 25; Mismatches 75; Indels 125; Gaps 12;

QY 12 YMSGTVNGHVEVEGDK-----GKPYEGEOTVRLVTKG-----GPLPFANDI-----LSPQ 59
 Db 2080 YYAGTTGGWRVHVHFKHLGC-----SVRFWYQGYPPAGSQVTVWALDNVYIGPQ 2132

QY 60 CQ---YG-----SIPFTKYPEDIPYVKQSFPGRYTWERIMNF----- 94
 Db 2133 CEEMCYHGSCINGTKICIDPGYSGPTCKISTKNPDLKDDFEGQLSDRFLMWSGGKPS 2192

QY 95 -----EDGAVCTVNSDSSIOGNCFIYHVKE-----SGLNFPNPGP 129
 Db 2193 RKCGILLSSGNLFFNEDGLRMLVTRDLDS-----HARFVQFMRLGCGKGVDPDRSQP 2246

QY 130 VM-----QKKTQGWEPNTERLF--- 146
 Db 2247 VLIQSYLSNGLSLSLQEFILFSSNVGRVIALEMPKARSGSTRLRWQPSENGHPYSP 2306

QY 147 -ARDGMLIGNFMALKLEGGHVLCPEKFKYAKKPKVMPG 186
 Db 2307 WVIDQILIGNI-----SGNTVLEDDFSTLDSRKLHFG 2341

RESULT 9
 SYA STAAW
 ID SYA STAAW STANDARD; PRT; 876 AA.
 AC Q9TNI;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
 GN ALAS OR SAV1618 OR SA1446.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878, 159879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mu50 / ATCC 700699, and N315;
 RX MEDLINE=21311952; PubMed=11478146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino J.-Q., Ito T.,
 RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RL Lancet 357:1225-1240 (2001).
 CC -|- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
 CC diphosphate + L-alanyl-tRNA(Ala).
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -|- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP003362; BAB57780.1; -
 CC EMBL; AP003364; BAB42710.1; -
 CC PIR; A89944; A89944.
 CC HAMAP; MF 00036; -; 1.
 CC InterPro; IPR003156; DHHA1.
 CC InterPro; IPR002318; tRNA-synt 2c.
 CC InterPro; IPR006193; tRNA_synt_Ala.
 CC Pfam; PF02272; DHHA1; 1.
 CC Pfam; PF01411; tRNA-synt 2c; 1.
 CC PRINTS; PR00980; TRNASYNTHALA.
 CC TIGRFAMs; TIGR00344; alas; 1.
 CC PROSITE; PS00860; AA tRNA LIGASE II ALA; 1.
 CC Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
 CC complete proteome.
 CC SQ SEQUENCE 876 AA; 98538 MW; 2B2BC79041AC264F CRC64;

Query Match 6.2%; Score 79.5; DB 1; Length 876;
 Best Local Similarity 21.4%; Pred. No. 24;
 Matches 43; Conservative 25; Mismatches 56; Indels 77; Gaps 10;

QY 16 TVNGH-YFEVEGKGKPYEGEOTVRLVTKGGLPFPADILSPQCYGSIP-----FTKY 70
 Db 81 TARHHTFFEMGLNFSIGDYFKQEAIE-----FAWEFLTSDKWMGMEPKLYVTIH 130

QY 71 PEDIPYVKQSFPGRYTWERIMNFEDGAVCTVNSDSSIOGNCFIYHVKFSGLNFPNPGP 130
 Db 131 PEDMEAY-----NIWHKDIGLESRII-----RIEGN-----FWDIGEGPSG-- 167

QY 131 MQKKTQGWEPNTERLFARDQMLIGNFMALKLEGGHVLCPEKFKYAKKPKVMPGYHYV 190
 Db 168 -----PNTLFIYDPR-GEAYGQDDPAEMYPGGEN----- 195

QY 191 DRKLDV-----TNHKNQY 204
 Db 196 ERYLEVNNLVFSEFNHNDHS 216

RESULT 10
 SYA STAAW
 ID SYA STAAW STANDARD; PRT; 876 AA.
 AC Q9NW87;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
 GN ALAS OR MW1568.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
 RA Yamanoto K., Hiramatsu K.;
 RT "Genome and virulence determinants of high virulence community-
 RT acquired MRSA";
 RL Lancet 359:1819-1827 (2002).
 CC -|- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
 CC diphosphate + L-alanyl-tRNA(Ala).
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -|- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP004827; BAB95433.1; -.
DR HANAP; MF 00036; 7; 1
DR InterPro; IPR003156; DHHA1.
DR InterPro; IPR002318; tRNA-synt_2c.
DR InterPro; IPR006193; tRNA_synt_Ala.
DR Pfam; PF02272; DHHA1; 1.
DR Pfam; PF01411; tRNA-synt_2c; 1.
DR PRINTS; PR00980; TRNASYNTHALA.
DR TIGRFAMs; TIGR00344; alas; 1.
DR PROSITE; PS50860; AA tRNA_LIGASE II ALA; 1.
KW Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 876 AA; 98505 MW; 5D9D662D8D8DAEDFC CRC64;

Query Match 6.2%; Score 79.5; DB 1; Length 876;
Best Local Similarity 21.4%; Pred. No. 24;
Matches 43; Conservative 25; Mismatches 56; Indels 77; Gaps 10;

QY 16 TWNGH-YEEVEGDGKPYEGEQTURLAVTKGGLPFAWDILSPQCQYGSIP----FTKY 70
DB 81 TARHHTFFEMGLNFSIGYFQKEAIE-----FAWEFLTSKWMGMEPDKLYVTIH 130

QY 71 PEDIPDYKQFPFGRYTWERIMNPFEDGAVCTVSNDSIIQGNCFIYHVKFSGLNFPNGPV 130
DB 131 PEDMEAY-----NIWKDGLBESRII-----RIEGN-----FWDIGEGPSG-- 167

QY 131 MQKKTQGWEPNTERLPARDGLMIGNFNFWALKLEGGGHVLCPEKSYKAKPKVMPGYHYV 190
DB 168 -----PNTETFYDNR-GEAYGDDPAEEMYPGGEN----- 195

QY 191 DRKLDV-----TNHNKDYT 204
DB 196 ERYLEVNLVVFSEFNHNKDS 216

RESULT 11
H136 ARATH
ID H136 ARATH STANDARD; PRT; 403 AA.
AC 082660;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Photosystem II stability/assembly factor HCF136, chloroplast precursor
DE HCF136 OR AT5G23120 OR MYJ24.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Wassilewskija; TISSUE=Leaf;
RX MEDLINE=98409534; PubMed=9736608;
RA Meurer J., Pluetschen H., Kowallik K.V., Westhoff P.;
RT "A nuclear-encoded protein of prokaryotic origin is essential for the
FT stability of photosystem II in Arabidopsis thaliana.";
RL EMOB J. 17:5286-5297(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20181125; PubMed=10718197;
RA Korani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RL DNA Res. 4:291-300(1997).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

```

```

RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Pale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Giang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";
RL Science 302:842-846(2003).
CC -!- FUNCTION: Essential for photosystem II (PSII) biogenesis; putative
CC protein stability or assembly factor for PSII.
CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen but
CC attached to the membrane. Restricted to the stromal lamellae.
CC -!- TISSUE SPECIFICITY: Expression in green tissue, not roots.
CC -!- DEVELOPMENTAL STAGE: Accumulates also in dark-grown seedlings.
CC -!- SIMILARITY: Belongs to the ycf48 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y15628; CAAT5723.1; -.
CC EMBL; AB006708; BAB09829.1; -.
CC EMBL; AY045691; AAK74049.1; -.
CC PIR; T51828; T51828.
CC DR InterPro; IPR002860; GH_BNR.
CC DR InterPro; IPR006311; Tat.
CC DR Pfam; PF02012; BNR; 4.
CC DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
KW Chloroplast; Thylakoid; Transit peptide; Photosystem II; Membrane.
FT TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
FT TRANSIT 54 78 THYLAKOID (POTENTIAL).
FT CHAIN 79 403 PHOTOSYSTEM II STABILITY/ASSEMBLY FACTOR
FT HCF136.
SQ SEQUENCE 403 AA; 44103 MW; 11079552F817FF9D CRC64;

Query Match 6.1%; Score 79; DB 1; Length 403;
Best Local Similarity 23.9%; Pred. No. 11;
Matches 49; Conservative 25; Mismatches 89; Indels 42; Gaps 11;

QY 18 NGHYPEVEGDGKPYEGEQTURLAVTKGGLPFAWDILSPQCQY-GSIPFTKYPED--- 73
DB 144 NYRNFNISFKGKEGWIIGKPAILLYTADAGE---NWDRIPLSSQLPGDMVFKATEDKSA 200

QY 74 --IPD----YUKQSPFGRYTWERIMNPFEDGAVCTVSNDSIIQGNCFIYHVKFSGLNFPNP 127
DB 201 EMTVDEGAIVTSN--RGYNWKAAIQETVSNATLARTVSSGISGASY-YTGTFSAVNRSPD 257

QY 128 GPVMQKKTQG-----WEPNT-----ERLFADGMLIG-----NFMALKLEGGGHVLC--- 170
DB 258 GRVAVSSRGNFFLTWEPGQYWPQHNRAVARRIQNWGWADGGIWLIVRGGLYLSXGT 317

QY 171 ----EFKSTYKAKPKVMPGYHYVD 191
DB 318 GITEEFEV-----FVQSRGFGILD 337

RESULT 12
CYBC BRAJA
ID CYBC BRAJA
AC P51131;
DT 01-OCT-1996 (Rel. 34, Created)

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b/c1 [includes: Cytochrome b; Cytochrome c1].
 GN FECH OR BLR486.
 OS Bradyrhizobium japonicum.
 OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=100R1P15;
 RA MEDLINE=89249332; PubMed=2541921;
 RA Thoeny-Meyer L., Stax D., Hennecke H.;
 RT "An unusual gene cluster for the cytochrome bcl complex in
 RT Bradyrhizobium japonicum and its requirement for effective root
 RT nodule symbiosis";
 RL Cell 57:683-697(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyumi T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110";
 RL DNA Res. 9:189-197(2002).
 CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 CC complex (complex III or cytochrome b-c1 complex), which is a
 CC respiratory chain that generates an electrochemical potential
 CC coupled to ATP synthesis. c1 functions as an electron donor to
 CC cytochrome c.
 CC -!- COFACTOR: Binds three hemes groups, two non-covalently and one
 CC covalently. Heme 1 (or BL or b562) is low-potential and absorbs at
 CC about 562, and heme 2 (or BH or b566) is high-potential and
 CC absorbs at about 566. Heme 3 is covalently bound like cytochrome
 CC c1.
 CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 CC cytochrome c1 and the Rieske protein.
 CC -!- PTM: The protein may be posttranslationally processed into
 CC cytochrome c1 and b.
 CC -!- SIMILARITY: Belongs to the cytochrome b family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J03176; AAA26200.1; -;
 DR EMBL; AF005944; BAC47751.1; -;
 DR PIR; B32382; B32382.
 DR InterPro; IPR002326; Cyt C1.
 DR InterPro; IPR005798; Cytb_b6_C.
 DR InterPro; IPR005797; Cytb_b6_N.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR Pfam; PF00032; cytochrome_b_C; 1.
 DR Pfam; PF00033; cytochrome_b_N; 1.
 DR Pfam; PF02167; Cytochrome c1; 1.
 DR PRINTS; PR00603; CYTOCHROME_C1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_QO; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Electron transport; Respiratory chain; Heme; Transmembrane;
 KW Complete proteome.
 RN DOMAIN 1 413 CYTOCHROME B.
 FT DOMAIN 414 687 CYTOCHROME C1.
 FT METAL 96 96 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 110 110 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 197 197 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL

FT METAL 211 211 IRON 2 (HEME B566 AXIAL LIGAND).
 FT BINDING 471 471 HEME (COVALENT).
 FT BINDING 474 474 HEME (COVALENT).
 FT METAL 475 475 IRON (HEME AXIAL LIGAND).
 FT METAL 616 616 IRON (HEME AXIAL LIGAND). (BY SIMILARITY).
 FT TRANSMEM 46 66 POTENTIAL.
 FT TRANSMEM 126 146 POTENTIAL.
 FT TRANSMEM 160 180 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 FT TRANSMEM 247 267 POTENTIAL.
 FT TRANSMEM 305 325 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 410 430 POTENTIAL.
 FT TRANSMEM 666 678 ANCHORS TO THE MEMBRANE (POTENTIAL).
 SQ SEQUENCE 687 AA; 76469 MW; A58048F1CF1EFFBB CRC64;
 Query Match 6.1%; Score 79; DB 1; Length 687;
 Best Local Similarity 22.1%; Pred. No. 20;
 Matches 40; Conservative 20; Mismatches 61; Indels 60; Gaps 7;
 Qy 2 VIAQMITYKVMSTGVNHYFEVGG---DGKGYEGEQTAVRLAVTKGGLPFAWDILSP 58
 Db 498 VAAFPASYKIKDGNPDAGDMFERGPRADYFPSPFNEQARAA--NGGAAPPDLSLITK 555
 Qy 59 QCQYXG-SIP-----FTKYPEDIPDYV-----KQSPFG----- 84
 Db 556 ARSYGRGFPWFIPDFFTQYQEGDPYVSALVQGEKVPGEVTPGSEYVKNKYPFGHAIK 615
 Qy 85 --RTWRIMNFDGAVCTVSNDS-----IQNCFTYHVKESGLN 123
 Db 616 MPKPLSDGQVTDYDDGSPATVAQYKDVTFMLMTAEHPMEARKLGFQVFLIFAGLM 675
 Qy 124 F 124
 Db 676 Y 676
 RESULT 13
 DPOL BPT4
 ID DPOL BPT4 STANDARD; PRT; 898 AA.
 AC P04415;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA polymerase (EC 2.7.7.7) (Gp43).
 GN 43.
 OS Bacteriophage T4.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC T4-like viruses.
 OX NCBI_TaxID=10665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88227938; PubMed=3286635;
 RA Spicer E.K., Rush J., Fung C., Reha-Krantz L.J., Karam J.D.,
 RA Konigsberg W.H.;
 RT "Primary structure of T4 DNA polymerase. Evolutionary relatedness to
 RT eucaryotic and other procaryotic DNA polymerases";
 RL J. Biol. Chem. 263:7478-7486(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89042102; PubMed=3054876;
 RA Andrade M., Guild N., Hsu T., Gold L., Tuerk C., Karam J.;
 RT "DNA polymerase of bacteriophage T4 is an autogenous translational
 RT repressor";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7942-7946(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22514363; PubMed=12626685;
 RA Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Ruger W.;
 RT "Bacteriophage T4 genome";
 RL Microbiol. Mol. Biol. Rev. 67:86-156(2003).
 RN [4]

SEQUENCE OF 338-898 FROM N.A.
 MEDLINE=88166734; PubMed=3350013;
 Laun N., Wang Y., Mathews C.K., Rueger W.;
 "Deoxycytidylate hydroxymethylase gene of bacteriophage T4.
 Nucleotide sequence determination and over-expression of the gene.";
 Eur. J. Biochem. 172:553-563(1988).
 [5]
 X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1-388.
 MEDLINE=96292335; PubMed=8679562;
 Wang J., Yu P., Lin T.C., Konigsberg W.H., Steitz T.A.;
 "Crystal structures of an NH2-terminal fragment of T4 DNA polymerase
 and its complexes with single-stranded DNA and with divalent metal
 ions.";
 Biochemistry 35:8110-8119(1996).
 -!- FUNCTION: This polymerase possesses two enzymatic activities: DNA
 synthesis (polymerase) and an exonucleolytic activity that
 degrades single stranded DNA in the 3' to 5' direction.
 -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 + {DNA} (N).
 -!- SIMILARITY: Belongs to the DNA polymerase type-B family.
 -!- DATABASE: NAME=Worthington enzyme manual;
 WWW="http://www.worthington-biochem.com/DNAPT4/".

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M10160; AAC05397.1; -;
 EMBL; X00769; CAA25344.1; -;
 EMBL; AF158101; AAD24268.1; -;
 EMBL; M37159; AAR21706.1; -;
 PIR; JS0791; DJBPT4.
 PDB; INOY; 14-OCT-96.
 PDB; INOZ; 14-OCT-96.
 InterPro; IPR006172; DNA pol B.
 InterPro; IPR006134; DNA pol_B_dom.
 InterPro; IPR006133; DNA pol_B_exo.
 Pfam; PF00136; DNA pol B; 1.
 Pfam; PF03104; DNA pol B_exo; 1.
 PRINTS; PR00106; DNA POLB.
 SMART; SM00486; POLBc; 1.
 PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 Transferrase; DNA-directed DNA polymerase; DNA replication; Hydrolase;
 Exonuclease; DNA-binding; 3D-structure.
 FT CONFLICT 89 89 A -> V (IN REF. 2; CAA25344).
 FT STRAND 4 11
 FT TURN 12 13
 FT STRAND 14 21
 FT TURN 22 24
 FT STRAND 25 31
 FT TURN 37 38
 FT STRAND 40 42
 FT HELIX 58 59
 FT STRAND 63 67
 FT HELIX 68 71
 FT TURN 83 83
 FT TURN 84 94
 FT HELIX 103 105
 FT STRAND 108 114
 FT TURN 123 125
 FT STRAND 132 137
 FT TURN 142 148
 FT STRAND 149 150
 FT TURN 152 153
 FT TURN 161 165
 FT TURN 168 171
 FT TURN 177 181
 FT TURN 182 182

183 188
 FT STRAND 191 204
 FT HELIX 209 211
 FT TURN 215 218
 FT STRAND 219 229
 FT HELIX 230 231
 FT TURN 233 235
 FT STRAND 236 238
 FT TURN 241 242
 FT STRAND 245 248
 FT STRAND 259 262
 FT TURN 263 264
 FT STRAND 266 267
 FT HELIX 270 277
 FT HELIX 287 295
 FT HELIX 307 313
 FT TURN 314 315
 FT STRAND 316 335
 FT HELIX 337 348
 FT TURN 349 349
 FT HELIX 352 356
 FT HELIX 358 368
 FT TURN 369 369
 SQ SEQUENCE 898 AA; 103609 MW; 925300C4CA5C7A24 CRC64;
 Query Match 6.1%; Score 79; DB 1; Length 898;
 Best Local Similarity 22.3%; Pred. No. 27;
 Matches 49; Conservative 37; Mismatches 66; Indels 68; Gaps 14;
 QY 41 LAVTKGGPLPFAWDILSPQCQYGSIPFTKYPED---IP---DYVKQSPFGRYTWE----- 89
 DB 342 LMSYVYAKMPFS-GVMSPIKTDWALIFNSLKGHEKVIQQGSHVKQSPFGAFVFPKPIA 400
 QY 90 --RMNPE-----DGAVCTVS-NDSSIOGNCFTY--HVKPSG-----LNFPNGPVM 131
 DB 401 RRYIMSFDTLSLPSIIRQVNIISPETIRGQPKVHPHEVIAGTAPKPSDEYSCSPNGWY 460
 QY 132 QKKTQGWEPN-----TERLFARDGLIGNNFMALK---LEGGHYLCEFKSTY 176
 DB 461 DRHQEGIIKPKIAKFFQKDWKKKMFABE-----MNAEAIKKIMKGAGS--CSTKPEV 513
 QY 177 KAKPKVPMGPYHYVDKLDVTNNKDYTS-----VEQCE 210
 DB 514 E-----RYVKFSDDFLNLSNYTESVLSLIECE 543
 RESULT 14
 ID LOXC ARATH STANDARD; PRT; 896 AA.
 AC P38418; OSW4E4; Q9MIU5;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Lipoxigenase, chloroplast precursor (EC 1.13.11.12).
 OS LOX2 OR AT3G45140 OR T14D3.80.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94120003; PubMed=8290626;
 RA Bell E., Mullet J.E.;
 RT "Characterization of an Arabidopsis lipoxigenase gene responsive to
 methyl jasmonate and wounding.";
 RL Plant Physiol. 103:1133-1137(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
 RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,

RA Dalseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choine N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erffle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Schaefer M., Schoen O., Bagues M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masny D.,
RA De Haan M., Maarsee A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirio A., Flores M., Liguri R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Malti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltischer J., Sellers P., Gill J.B., Feidblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gujral M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Xuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";
RL Science 302:842-846(2003).
CC -!- FUNCTION: PLANT LIPOXYGENASE MAY BE INVOLVED IN A NUMBER OF
CC DIVERSE ASPECTS OF PLANT PHYSIOLOGY INCLUDING GROWTH AND
CC DEVELOPMENT, PEST RESISTANCE, AND SENESENCE OR RESPONSES TO
CC WOUNDING. IT CATALYZES THE HYDROPEROXIDATION OF LIPIDS, CONTAINING
CC A CIS, CIS-1,4-PENTADIENE STRUCTURE.
CC -!- CATALYTIC ACTIVITY: Linoleate + O(2) = (9Z,11E)-(13S)-13-
CC hydroperoxyoctadeca-9,11-dienoate.
CC -!- COFACTOR: Iron, one atom tightly bound per molecule (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast (Potential).
CC -!- TISSUE SPECIFICITY: In leaves and inflorescences but not abundant
CC in seeds, roots and stems.
CC -!- INDUCTION: By methyl jasmonate and wounding.
CC -!- SIMILARITY: Belongs to the lipoxygenase family.
CC -!- SIMILARITY: Contains 1 Pfam domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L23968; AAA32749.1; -.

DR EMBL; ALI38649; CAB72152.1; ALT_SEQ.
DR EMBL; AY062611; AAL32689.1; -.
DR PIR; JQ2391; JQ2391.
DR HSP; P08170; SBL.
DR SWISS-2DPAGE; P38418; ARATH.
DR InterPro; IPR000907; lipoxygenase.
DR InterPro; IPR01024; lipoxygenase_LH2.
DR InterPro; IPR008976; PLAT LH2.
DR Pfam; PF00305; lipoxygenase; 1.
DR Pfam; PF01477; PLAT; 1.
DR PRINTS; PR00087; LIPOXYGENASE.
DR SMART; SM00308; LH2; 1.
DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
DR PROSITE; PS00095; PLAT; 1.
DR Oxidoreductase; Dioxygenase; Iron; Multigene family; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 896 LIPOXYGENASE.
FT DOMAIN 79 199 PLAT.
FT METAL 554 554 IRON (BY SIMILARITY).
FT METAL 559 559 IRON (BY SIMILARITY).
FT METAL 746 746 IRON (BY SIMILARITY).
FT METAL 896 896 IRON (BY SIMILARITY).
FT CONFLICT 613 613 W -> C (IN REF. 3).
SQ SEQUENCE 896 AA; 102045 MW; F32822205C8F9F22 CRC64;
Query Match 6.1%; Score 78.5; DB 1; Length 896;
Best Local Similarity 21.0%; Pred No. 30;
Matches 57; Conservative 31; Mismatches 97; Indels 87; Gaps 12;
Qy 9 YKVMSTGVNGHYFEVGDG-----KGRPYE---GEQTVRLAVTKGGPLPFAWDILSP 58
Db 617 VALSLSAVYKGLWRFDQGLPADLKRGLAEEDKTAHGVRLTIP---DYFPAND--- 669
Qy 59 QCQVGSIPFTKYPEDIPYKQSPFGTYWIRMNFDGAVCTVSNDSISQNCFIYHK 118
Db 670 ----GLILWDAIKEWVTDYVKHYYPDEE-----LITSDEELQG-----W 704
Qy 119 FSLGNLFPNGFVGMOKTQGMENPTELFARDGMLGNFMALKEGGGHYLCEP---KS 174
Db 705 WSEVRNIGHG---DKDEPWWP-----VLKTDQDILGVTTIAWYTSGHAAVNGQYCYG 757
Qy 175 TYKAKXP-----VKMPGYHYVDKLD-----VTNNHK 201
Db 758 GYFENRPTTRIRMTEDPTDEALKEFYESPEKVLKLTYPKQKQATLWMVTLDLLSTHP 817
Qy 202 DYTSV-ROCELSIARKPVACRFPRVKSRYK 232
Db 818 DEEYIGEQEASWANEPVINAAPERFKGLQY 849
RESULT 15
RELN_RAT
ID RELN_RAT STANDARD; PRT; 3462 AA.
AC P58751; Q80T65;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reelin precursor (EC 3.4.21.-).
GN RELN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Cerebellum;
RA Kikkawa S., Terashima T.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND DISEASE.
RC TISSUE=Cerebellum;

KW		Alternative splicing.		POTENTIAL.	
FT	SIGNAL	1	27		
FT	CHAIN	28	3462		REELIN.
FT	DOMAIN	41	173		REELER.
FT	DOMAIN	672	703		EGF-LIKE 1.
FT	DOMAIN	1031	1062		EGF-LIKE 2.
FT	DOMAIN	1410	1443		EGF-LIKE 3.
FT	DOMAIN	1766	1797		EGF-LIKE 4.
FT	DOMAIN	2130	2162		EGF-LIKE 5.
FT	DOMAIN	2479	2510		EGF-LIKE 6.
FT	DOMAIN	2854	2885		EGF-LIKE 7.
FT	DOMAIN	3229	3261		EGF-LIKE 8.
FT	REPEAT	594	605		BNR 1.
FT	REPEAT	800	811		BNR 2.
FT	REPEAT	953	964		BNR 3.
FT	REPEAT	1158	1169		BNR 4.
FT	REPEAT	1324	1335		BNR 5.
FT	REPEAT	1536	1547		BNR 6.
FT	REPEAT	1687	1698		BNR 7.
FT	REPEAT	1885	1896		BNR 8.
FT	REPEAT	2044	2055		BNR 9.
FT	REPEAT	2251	2262		BNR 10.
FT	REPEAT	2400	2411		BNR 11.
FT	REPEAT	2599	2610		BNR 12.
FT	REPEAT	2779	2790		BNR 13.
FT	REPEAT	2980	2991		BNR 14.
FT	REPEAT	3364	3375		BNR 15.
FT	DOMAIN	3433	3462		ARG-RICH (BASIC).
FT	CARBOHYD	142	142		N-LINKED (GLCNAC. .)
FT	CARBOHYD	259	259		N-LINKED (GLCNAC. .)
FT	CARBOHYD	291	291		N-LINKED (GLCNAC. .)
FT	CARBOHYD	307	307		N-LINKED (GLCNAC. .)
FT	CARBOHYD	630	630		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1268	1268		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1448	1448		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1601	1601		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1751	1751		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1922	1922		N-LINKED (GLCNAC. .)
FT	CARBOHYD	2146	2146		N-LINKED (GLCNAC. .)
FT	CARBOHYD	2270	2270		N-LINKED (GLCNAC. .)
FT	CARBOHYD	2318	2318		N-LINKED (GLCNAC. .)
FT	CARBOHYD	2570	2570		N-LINKED (GLCNAC. .)
FT	CARBOHYD	2963	2963		N-LINKED (GLCNAC. .)
FT	CARBOHYD	3017	3017		N-LINKED (GLCNAC. .)
FT	CARBOHYD	3074	3074		N-LINKED (GLCNAC. .)
FT	CARBOHYD	3186	3186		N-LINKED (GLCNAC. .)
FT	CARBOHYD	3413	3413		N-LINKED (GLCNAC. .)
FT	CARBOHYD	3440	3440		N-LINKED (GLCNAC. .)
FT	VARSPPLIC	3430	3431		Missing (in isoform 2).
FT	FTId=VSP	005579.			/FTId=VSP 005579.
FT	VARSPPLIC	3430	3462		Missing (in isoform 3).
FT	FTId=VSP	005580.			/FTId=VSP 005580.
FT	CONFLICT	336	336		H -> R (IN REF. 2).
FT	CONFLICT	2714	2714		V -> L (IN REF. 2).
FT	SEQUENCE	3462 AA;	387525 MW;	FCCF89B090E035F6	CRC64;
Query Match 6.1%; Score 78.5; DB 1; Length 3462;					
Best Local Similarity 19.6%; Pred.No.1.5e+02;					
Matches 55; Conservative 26; Mismatches 75; Indels 125; Gaps 12;					
QY	12	YMSGTVNGHYFVEVGDGK	-----GKPYEGEQTVRLAVTKG	-----GPIFFAWDI	-----LSPO 59
Db	2081	YYAGTIGQWRREWHFGKHLG	-----SVFRWYQGFYFAGSQPVTVADNVYIGQ	-----	2133
QY	60	CQ	-----YQSI	-----PFTKYPEDIPDYVKQSPFGRYTWBRIMNF	-----94
Db	2134	CEMCCGHCSCVNGTKICDPGVSGETCKISTKNPDLKDDPEGQLESDFILMSGKPS	-----	-----	2193
QY	95	-----	-----EDGAVCTVSNDSSITQNCFIYHKVF	-----	-----SGINFPNCP 129
Db	2194	RKCGILSSGNNLFFNEBDGLRMVLTRDLDS	-----HARFVQFFMRLGCGKGVDPFRSQP	-----	2247
QY	130	VM	-----	-----OKKTOGEPNTERIF	-----146

Db 2248 VLLQYSLNGGLSWELLQEFLLFSNSSNVGRYIALEMLKARSGSTRLEWOPSENGHFYSP 2307
QY 147 -ARDGMLIGNNFALKLEGGHYLCEFKSTYKAKPVKMPG 186
Db 2308 WVIDQILIGGNI-----SGNTVLEDDFSTLDSRKLHHPG 2342

Search completed: August 12, 2004, 06:20:09
Job time : 15.4467 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:19:43 ; Search time 494.559 Seconds
(without alignments)
149.169 Million cell updates/sec

Title: US-09-890-463-4

Perfect score: 1287

Sequence: 1 SVIAKQMTYKVMSTGNH.....KPVVACRFVRKSRKHYAVA 235

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	788	61.2	225	15	US-10-442-148A-7
2	788	61.2	239	15	US-10-442-148A-8
3	785	61.0	225	14	US-10-315-920-6
4	784	60.9	225	14	US-10-121-258-20
5	784	60.9	225	14	US-10-315-920-4
6	784	60.9	236	16	US-10-314-936-2
7	784	60.9	236	16	US-10-314-936-4
8	783	60.8	225	9	US-09-999-745-67
9	783	60.8	225	10	US-09-866-538-12
10	783	60.8	225	10	US-09-794-308-12
11	783	60.8	225	10	US-09-865-291-12
12	783	60.8	225	12	US-10-132-067-4
13	783	60.8	225	13	US-10-006-922-12
14	783	60.8	225	14	US-10-081-864-8
15	783	60.8	225	14	US-10-121-258-1

16	783	60.8	225	14	US-10-315-920-2
17	783	60.8	225	15	US-10-370-570-56
18	783	60.8	225	15	US-10-406-618-32
19	783	60.8	225	16	US-10-433-640-13
20	783	60.8	240	14	US-10-152-296-2
21	783	60.8	240	16	US-10-739-656-2
22	783	60.8	487	15	US-10-343-977-1
23	783	60.8	506	15	US-10-343-977-2
24	783	60.8	545	14	US-10-214-932-52
25	783	60.8	547	15	US-10-343-977-3
26	783	60.8	548	14	US-10-214-932-76
27	778	60.5	225	16	US-10-423-688A-41
28	773	60.1	227	13	US-10-006-922-10
29	769	59.8	225	14	US-10-121-258-24
30	769	59.8	226	14	US-10-121-258-6
31	768	59.7	225	13	US-10-006-922-44
32	768	59.7	225	14	US-10-081-864-12
33	768	59.7	225	14	US-10-121-258-4
34	764.5	59.4	226	16	US-10-724-178-12
35	750	58.3	230	13	US-10-006-922-18
36	750	58.3	230	14	US-10-161-403-40
37	737.5	57.3	232	14	US-10-133-973-5
38	737.5	57.3	232	15	US-10-370-570-64
39	729.5	56.7	232	16	US-10-724-178-11
40	717	55.7	225	14	US-10-121-258-8
41	717	55.7	225	16	US-10-724-178-16
42	706	54.9	225	15	US-10-370-570-61
43	689	53.5	205	13	US-10-006-922-46
44	671	52.1	231	14	US-10-133-973-6
45	671	52.1	266	13	US-10-006-922-4

ALIGNMENTS

RESULT 1
US-10-442-148A-7
; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO
; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583-0103-00000
; CURRENT APPLICATION NUMBER: US/10/442.148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

Query Match	61.2%	Score 788;	DB 15;	Length 225;
Best Local Similarity	63.0%;	Pred. No. 4.3e-76;		
Matches 145;	Conservative 27;	Mismatches 44;	Indels 14;	Gaps 1;
Qy	1	SVIAKQMTYKVMSTGNHGFVEGDKGKPYEGEQTAVRLAVTKGGFLPFPAWILSPQC	60	
Db	6	NVITEFMRFKVRMBGTGNGHEFEIEGEGEPYEGHNTKLVKTKGGFLPFPAWILSPQF	65	
Qy	61	QYGSIFPTKYPEDIPDYVKSGFPGRYTWIRIMNPEDGAVCTVSNDSIQGNCFTYHVKFS	120	
Db	66	QYGSKVYKHPADIPDYKLSFPFGKWERVMNPDGGVATVTDSSLDQDCFTYKVKFI	125	
Qy	121	GLNPPPGVWQKKTQGWENPTERLFPARDGMLGNFMALKLEGHLYLCEFKSTYKAKK	180	


```
RESULT 5
US-10-315-920-4
; Sequence 4, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-4

Query Match          60.9%; Score 784; DB 14; Length 225;
Best Local Similarity 63.0%; Pred. No. 1.2e-75;
Matches 145; Conservative 27; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAKQMTYKYVMSTGTVNGHYFEVGGDKGKPYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 6 NVIKFMRKVRMEGTVMGHEFEIEGEGRPYEGHNTVKLVKTKGGPLPFAWDILSPQF 65

QY 61 QYGSIPFTKYPEDIPDYVKQSPGRYTWERIMNFEDGAVCTVSNDSSTQGNCFIYHVKFS 120
   ||||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 66 QYGSKVYVKHPADIPDYKKLSFPEGKWERVMNFEDGGVVTVTQDSSLQDGCIFYKVKFI 125

QY 121 GLNFPNGPVNQKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKAKK 180
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 126 GVNFPDGPVNMQKTMGWEASTERLYPRDGVLGKEIHKALKKLDGGHYLVFEKTIYMAKK 185

QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQCEISIAKPKVPVACRFRVRSRH 230
   ||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 186 PVQLPGYYVDSKLDITSHNKDYTIIVEQYE-----RTEGRH 221

RESULT 6
US-10-314-936-2
; Sequence 2, Application US/10314936
; Publication No. US20040110225A1
; GENERAL INFORMATION:
; APPLICANT: Gibbs, Patrick D.L.
; APPLICANT: Carter, Robert W.
; APPLICANT: Schmale, Michael C.
; TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
; FILE REFERENCE: 638.004
; CURRENT APPLICATION NUMBER: US/10/314,936
; CURRENT FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: mutant red fluorescent protein
US-10-314-936-2

Query Match          60.9%; Score 784; DB 16; Length 236;
Best Local Similarity 63.0%; Pred. No. 1.2e-75;
Matches 145; Conservative 27; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAKQMTYKYVMSTGTVNGHYFEVGGDKGKPYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 6 NVIKFMRKVRMEGTVMGHEFEIEGEGRPYEGHNTVKLVKTKGGPLPFAWDILSPQF 65

QY 61 QYGSIPFTKYPEDIPDYVKQSPGRYTWERIMNFEDGAVCTVSNDSSTQGNCFIYHVKFS 120
   ||||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 66 QYGSKVYVKHPADIPDYKKLSFPEGKWERVMNFEDGGVVTVTQDSSLQDGCIFYKVKFI 125

QY 121 GLNFPNGPVNQKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKAKK 180
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 126 GVNFPDGPVNMQKTMGWEASTERLYPRDGVLGKEIHKALKKLDGGHYLVFEKTIYMAKK 185

QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQCEISIAKPKVPVACRFRVRSRH 230
   ||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 186 PVQLPGYYVDSKLDITSHNKDYTIIVEQYE-----RTEGRH 221

RESULT 7
US-10-314-936-4
; Sequence 4, Application US/10314936
; Publication No. US20040110225A1
; GENERAL INFORMATION:
; APPLICANT: Gibbs, Patrick D.L.
; APPLICANT: Carter, Robert W.
; APPLICANT: Schmale, Michael C.
; TITLE OF INVENTION: FLUORESCENT PROTEINS FROM AQUATIC SPECIES
; FILE REFERENCE: 638.004
; CURRENT APPLICATION NUMBER: US/10/314,936
; CURRENT FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: mutant red fluorescent protein
US-10-314-936-4

Query Match          60.9%; Score 784; DB 16; Length 236;
Best Local Similarity 63.0%; Pred. No. 1.2e-75;
Matches 145; Conservative 27; Mismatches 44; Indels 14; Gaps 1;

QY 1 SVIAKQMTYKYVMSTGTVNGHYFEVGGDKGKPYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 6 NVIKFMRKVRMEGTVMGHEFEIEGEGRPYEGHNTVKLVKTKGGPLPFAWDILSPQF 65

QY 61 QYGSIPFTKYPEDIPDYVKQSPGRYTWERIMNFEDGAVCTVSNDSSTQGNCFIYHVKFS 120
   ||||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 66 QYGSKVYVKHPADIPDYKKLSFPEGKWERVMNFEDGGVVTVTQDSSLQDGCIFYKVKFI 125

QY 121 GLNFPNGPVNQKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKAKK 180
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 126 GVNFPDGPVNMQKTMGWEASTERLYPRDGVLGKEIHKALKKLDGGHYLVFEKTIYMAKK 185

QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQCEISIAKPKVPVACRFRVRSRH 230
   ||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 186 PVQLPGYYVDSKLDITSHNKDYTIIVEQYE-----RTEGRH 221

RESULT 8
US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
```

```
QY 1 SVIAKQMTYKYVMSTGTVNGHYFEVGGDKGKPYEGEQTIVRLAVTKGGPLPFAWDILSPQC 60
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 6 NVIKFMRKVRMEGTVMGHEFEIEGEGRPYEGHNTVKLVKTKGGPLPFAWDILSPQF 65

QY 61 QYGSIPFTKYPEDIPDYVKQSPGRYTWERIMNFEDGAVCTVSNDSSTQGNCFIYHVKFS 120
   ||||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 66 QYGSKVYVKHPADIPDYKKLSFPEGKWERVMNFEDGGVVTVTQDSSLQDGCIFYKVKFI 125

QY 121 GLNFPNGPVNQKKTQGWEPNTERLFARDGMLIGNFMALKLEGGHVLCBFKSTYKAKK 180
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 126 GVNFPDGPVNMQKTMGWEASTERLYPRDGVLGKEIHKALKKLDGGHYLVFEKTIYMAKK 185

QY 181 PVKMPGHHYVDRKLDVTNNHKDYSVEQCEISIAKPKVPVACRFRVRSRH 230
   ||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 186 PVQLPGYYVDSKLDITSHNKDYTIIVEQYE-----RTEGRH 221

RESULT 9
US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
```


66	QYGSKVYVKKPADLPDYKKUSFPFGFKWERNVFNEDGGVTVTQDSSYQDGGFTYKVKPI	125
Db	: : : : : : : :	
121	GLNFPPNGPYWKKTQGWEPNTELFARFADGMLIGNNFMALKLEGGGHYLCFSTYKAKK	180
QY	: : : : : : :	
126	GVNFPSDGPYWKKTGMGWEASTERLYPRDGVLGKEIHKALKDGGHYLVFEKSYIMAKK	185
Db	: : : : : : :	
181	PVKMPGHHYVDRKLDVTNNKNDYTSVEQCEISIARKFPVACRFPVRGRH	230
QY	: : : : : : :	
186	PVQLPGYYVDSKLDITSHNEDYTIVEQYE-----RTEGRH	221
Db	: : : : : : :	

```

RESULT 15
US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Teien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(225)
; OTHER INFORMATION: wild-type DsRed
US-10-121-258-1

```

[illegible]

Search completed: August 12, 2004, 06:51:22
Job time : 495.559 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 24.5594 Seconds
(without alignment)
493.990 Million cell updates/sec

Title: US-09-890-463-4
Perfect score: 1287
Sequence: 1 SVIAKQMTYKVMSTGVNH.....KPVVACRFVRKSRKXAVA 235

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	479.5	37.3	238	3	US-09-277-716-16
2	479.5	37.3	238	4	US-09-609-161B-16
3	479.5	37.3	238	4	US-09-626-581D-65
4	479.5	37.3	238	4	US-09-415-765B-65
5	479.5	37.3	238	4	US-09-626-580C-65
6	470.5	36.6	238	3	US-09-277-716-32
7	470.5	36.6	238	4	US-09-609-161B-32
8	468.5	36.4	238	4	US-09-839-650-3
9	138.5	15.4	238	4	US-09-472-065A-5
10	133.5	15.0	238	4	US-09-479-645A-12
11	132.5	15.0	238	4	US-09-023-946B-28
12	130.5	14.8	238	4	US-09-023-946B-23
13	189.5	14.7	238	4	US-09-023-946B-36
14	189	14.7	238	4	US-09-023-946B-22
15	189	14.7	941	4	US-09-513-783A-172
16	188.5	14.6	238	1	US-08-337-915A-2
17	188.5	14.6	238	1	US-08-753-143-2
18	188.5	14.6	238	2	US-08-678-865-2
19	188.5	14.6	238	2	US-08-680-876-2
20	188.5	14.6	238	2	US-08-792-553-2
21	188.5	14.6	238	3	US-08-893-327-16
22	188.5	14.6	238	3	US-08-753-144-2
23	188.5	14.6	238	3	US-09-094-359-2
24	188.5	14.6	238	3	US-09-172-063-2
25	188.5	14.6	238	3	US-09-121-539-1
26	188.5	14.6	238	3	US-09-263-975-2
27	188.5	14.6	238	4	US-08-727-452-2

28	188.5	14.6	238	4	US-09-418-785-1	Sequence 1, Appli
29	188.5	14.6	238	4	US-09-214-909-2	Sequence 2, Appli
30	188.5	14.6	238	4	US-09-479-645A-10	Sequence 10, Appl
31	188.5	14.6	238	4	US-09-479-645A-159	Sequence 159, App
32	188.5	14.6	238	4	US-09-129-192C-2	Sequence 2, Appli
33	188.5	14.6	238	4	US-09-129-192C-74	Sequence 74, Appl
34	188.5	14.6	238	4	US-09-602-641-2	Sequence 2, Appli
35	188.5	14.6	238	4	US-09-704-463-2	Sequence 2, Appli
36	188.5	14.6	238	4	US-09-472-065A-4	Sequence 4, Appli
37	188.5	14.6	238	4	US-09-472-065A-6	Sequence 6, Appli
38	188.5	14.6	238	4	US-09-023-946B-29	Sequence 29, Appl
39	188.5	14.6	238	4	US-09-920-922-4	Sequence 4, Appli
40	188.5	14.6	238	5	PCT-US95-14692-2	Sequence 2, Appli
41	188.5	14.6	239	3	US-08-646-538-2	Sequence 2, Appli
42	188.5	14.6	239	3	US-09-094-359-4	Sequence 4, Appli
43	188.5	14.6	239	3	US-09-172-063-3	Sequence 3, Appli
44	188.5	14.6	239	3	US-09-503-222-2	Sequence 2, Appli
45	188.5	14.6	239	4	US-09-513-783A-46	Sequence 46, Appl

RESULT 1

US-09-277-716-16

; Sequence 16, Application US/09277716A

; Patent No. 6232107

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

; CURRENT APPLICATION NUMBER: US/09/277,716A

; EARLIER FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: 60/102,939

; EARLIER FILING DATE: 1998-10-01

; EARLIER APPLICATION NUMBER: 60/089,367

; EARLIER FILING DATE: 1998-06-15

; EARLIER APPLICATION NUMBER: 60/079,624

; EARLIER FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 16

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Renilla mulleri

; FEATURE:

; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Query Match 37.3%; Score 479.5; DB 3; Length 238;

Best Local Similarity 44.0%; Pred. No. 8.8e-45;

Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

Qy	7	MTYKVMGTVNGHYFVEGDKGKPYEGEQTAVLTGKGLPFPADWILSPQOYGSIP	66
Db	15	MSYKVNLEGINVNHVFTMEGCKGNILFGNLVQIRVTKGAPLPFADIVSPAFQYGNRT	74
Qy	67	FTKYPEDIPDYVKQSPFGRYTWERIMNPFEDGAVCTVNSDSSIQNCFTYHVKFSGLNPPP	126
Db	75	FTKYPNDISDYFIQSFPAGFMYERTLAVEDGLVEIRSDINLIEDKFVYRVEYKGSNFPD	134
Qy	127	NGPVMQKKTQWENTERTLRFARDGMLTGNPFMAIKLGGGHYLCFEPKSTYKAKPKVK-MP	185
Db	135	DGPVMQKTLTGIEFSFEAMNNNGVLGVILVYKLSKGKYSCMKTLKMSKGVKREFF	194
Qy	186	GHYHVDRLDVTNHNKDYTSVQCIEISAR-----KPV	218
Db	195	SYHFIQHRLEKT-YVEDGGFVEQHETALTAQWTSIGKPL	231

RESULT 2

US-09-609-161B-16

```
; Sequence 16, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR FILING DATE: 2000-07-27
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-16

Query Match      37.3%; Score 479.5; DB 4; Length 238;
Best Local Similarity 44.0%; Pred. No. 8.8e-45;
Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

Qy 7 MTKYVMGTVNGHYFVEVGGKPKYEGEQTVPVLAATKGGPLPFAWDILSPQCYGSIP 66
Db 15 MSYKVNLEGIVNNHVFTEGCGKNILFGNLQVQIRVTKGAPLFAFDIVSPAFQYGNRT 74

Qy 67 FTYPEDIPDYVKQSPFGRYTWERIMNFDGAVCTVNSDSIQNCFIYHVKFSGLNFP 126
Db 75 FTYPENDISDYFIQSFPAGFWERTLYEDGGLVEIRSDINLIEDKFVRYVEYKGSNFPD 134

Qy 127 NGPVMQKKTQGWEPNTERLFPARDGMLIGNNFMALKEGGHYLCFEKSTYKAKPKVK-MP 185
Db 135 DGPVMQKKTILGIEPSFEAMYNNGVLVGEVILVYKLSNGKYYSCHMTLMKSKGVVKEFP 194

Qy 186 GYHVVDRKLDVTNNKDYTSVEQCEISAR-----KPV 218
Db 195 SYHFIQHRLEKT-YVEDGGFVEQHEHTAIAQWTSIGKPL 231

RESULT 4
US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-09-415-765B-65

Query Match      37.3%; Score 479.5; DB 4; Length 238;
Best Local Similarity 44.0%; Pred. No. 8.8e-45;
Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

Qy 7 MTKYVMGTVNGHYFVEVGGKPKYEGEQTVPVLAATKGGPLPFAWDILSPQCYGSIP 66
Db 15 MSYKVNLEGIVNNHVFTEGCGKNILFGNLQVQIRVTKGAPLFAFDIVSPAFQYGNRT 74

Qy 67 FTYPEDIPDYVKQSPFGRYTWERIMNFDGAVCTVNSDSIQNCFIYHVKFSGLNFP 126
Db 75 FTYPENDISDYFIQSFPAGFWERTLYEDGGLVEIRSDINLIEDKFVRYVEYKGSNFPD 134

Qy 127 NGPVMQKKTQGWEPNTERLFPARDGMLIGNNFMALKEGGHYLCFEKSTYKAKPKVK-MP 185
Db 135 DGPVMQKKTILGIEPSFEAMYNNGVLVGEVILVYKLSNGKYYSCHMTLMKSKGVVKEFP 194

Qy 186 GYHVVDRKLDVTNNKDYTSVEQCEISAR-----KPV 218
Db 195 SYHFIQHRLEKT-YVEDGGFVEQHEHTAIAQWTSIGKPL 231

RESULT 3
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 2000-07-27
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
US-09-626-581D-65
```

```
Query Match      37.3%; Score 479.5; DB 4; Length 238;
Best Local Similarity 44.0%; Pred. No. 8.8e-45;
Matches 96; Conservative 41; Mismatches 74; Indels 7; Gaps 3;

Qy 7 MTKYVMGTVNGHYFVEVGGKPKYEGEQTVPVLAATKGGPLPFAWDILSPQCYGSIP 66
Db 15 MSYKVNLEGIVNNHVFTEGCGKNILFGNLQVQIRVTKGAPLFAFDIVSPAFQYGNRT 74

Qy 67 FTYPEDIPDYVKQSPFGRYTWERIMNFDGAVCTVNSDSIQNCFIYHVKFSGLNFP 126
Db 75 FTYPENDISDYFIQSFPAGFWERTLYEDGGLVEIRSDINLIEDKFVRYVEYKGSNFPD 134

Qy 127 NGPVMQKKTQGWEPNTERLFPARDGMLIGNNFMALKEGGHYLCFEKSTYKAKPKVK-MP 185
Db 135 DGPVMQKKTILGIEPSFEAMYNNGVLVGEVILVYKLSNGKYYSCHMTLMKSKGVVKEFP 194

Qy 186 GYHVVDRKLDVTNNKDYTSVEQCEISAR-----KPV 218
Db 195 SYHFIQHRLEKT-YVEDGGFVEQHEHTAIAQWTSIGKPL 231

RESULT 5
US-09-626-580C-65
; Sequence 65, Application US/09626580C
; Patent No. 6562617
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
```

```

; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US/09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US/09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-580C-65

```

Query Match	37.3%	Score 479.5	DB 4	Length 238
Best Local Similarity	44.0%	Pred. No. 8.8e-45		
Matches	96	Conservative 41	Mismatches 74	Indels 7
Gaps	3			
QY	7	MTYKVTMSGFVNGHYHFEVGGDGKPKYEGQTVRLAVTKGSPLPFAMDLILSPQOCYQSGIP	66	
Db	15	MSYKVNLEGIVNNHVFVMECCGKNILIFGNQLVOIRVTKGAPLPFADIVSFAFYQNGRT	74	
QY	67	FTKYPEDIPYVQKSPGRYTWIRIMNFEDGACVTVSNDSSIQNCFTYHYVKSGLNPPF	126	
Db	75	FTKYPNDISDYFIQSPFAGMYERTLAYEDGGLVEIRSDINLIBDKFVIRVYKSGNFPD	134	
QY	127	NGPVMOKTQGWBPNTLERLFAQDMLIGNPFMAKLGGGGHYLCEFKSTYKAKKPVK-MP	185	
Db	135	DGPVMQKTLIGIPEFSFAMYNWNGVLVGEVILVYKLSNGRYYSCHMKTILMKSQGVVKEFP	194	
QY	186	GYHYVRDKLDVTHNKNKYTSVEQCEISAR-----KPV	218	
Db	195	SVYFIIQHRLEKT-YVEDGGFVEOHEHTAIAOWTISICKPL	231	

```

RESULT 6
US-09-277-716-32
; Sequence 32, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277.716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-277-716-32

```

75	FTKYDDDIADYFVQSFPGAFYFYERNLPEDEGAIVDRSDISLEDDKFKHYKVFYRNGNGFPS	134
Db		
127	NGPVMKQTQGWENTERLFLARDGMLTGNFMALKEGGGHYLCFEPKSTTYAKPKVK-MP	185
QY	: : : : :	
Db	: : : : :	
135	NGPVMKAILGMEPSFEVYVNMNSGVLVGEVDLVYKLESGNYYSCHMKTFYRSRGGVKFP	194
Db	: : : : :	
186	GYYHYDRKLDVTNHNKNDYTSVEQCEIGIAR-----KEV	218
QY	: : : : :	
Db	: : : : :	
195	EYHFTHHRLKLT-KYVEEGSFVEQETHAIQLTITIGKPL	231
Db	: : : : :	

```

RESULT 7
US-09-609-161B-32
; Sequence 32, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS EN
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-609-161B-32

```

```

RESULT 8
US-09-839-650-3
; Sequence 3, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein

```


Qy	181	PVKMPGYHYVDRKLDVT---NHNKDY-----TsVE	207
Db	193	PVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKSGMSVVGIDLGFQ	252
Qy	208	QCEISIAR-----KPYVACRFFRVKSRHKYAVA	235
Db	253	SCYVAVARAGGIETIANEYSDRCTPACISFGPKNRSIGAAA	293

Search completed: August 12, 2004, 06:21:07
Job time : 25.5594 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:34:08 ; Search time 16 seconds
(without alignments)
30.060 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 206

Minimum DB seq length: 0

Maximum DB seq length: 5

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	47.6	4	2	PL0146
2	9	42.9	4	2	I40505
3	9	42.9	4	2	A27897
4	9	42.9	5	2	S11127
5	8	38.1	4	2	S17255
6	8	38.1	4	2	T30569
7	8	38.1	4	2	E44823
8	8	38.1	4	2	A60521
9	8	38.1	5	2	T14908
10	7	33.3	4	2	I61883
11	7	33.3	4	2	P70551
12	7	33.3	4	2	P70697
13	7	33.3	4	2	I37013
14	7	33.3	4	2	I84439
15	7	33.3	5	2	I39964
16	7	33.3	5	2	I39966
17	7	33.3	5	2	I39965
18	7	33.3	5	2	T14910
19	7	33.3	5	2	S55237
20	7	33.3	5	2	P70597
21	6	28.6	4	2	S18401
22	6	28.6	4	2	A61300
23	6	28.6	4	2	P70534
24	6	28.6	4	2	P70721
25	6	28.6	5	2	A44955
26	6	28.6	5	2	B61168
27	6	28.6	5	2	I50385
28	6	28.6	5	2	P70278
29	6	28.6	5	2	P70295

30	6	28.6	5	2	S62883
31	6	28.6	5	2	B44817
32	6	28.6	5	2	D44817
33	5	23.8	3	3	GRHU
34	5	23.8	3	3	A60898
35	5	23.8	3	3	S13894
36	5	23.8	3	3	E37196
37	5	23.8	3	3	F37196
38	5	23.8	3	3	F70578
39	5	23.8	4	2	A02147
40	5	23.8	4	2	A37832
41	5	23.8	4	2	I40870
42	5	23.8	4	2	I40804
43	5	23.8	4	2	T46627
44	5	23.8	4	2	JQ1273
45	5	23.8	4	2	S39390

ALIGNMENTS

RESULT 1

PL0146
Carbon-monoxide dehydrogenase (EC 1.2.99.2) small chain - Pseudomonas carboxydohydrogena
C:Species: Pseudomonas carboxydohydrogena
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 28-Apr-1993
C:Accession: PL0146
R:Kraut, M.; Hugendieck, I.; Herwig, S.; Meyer, O.
Arch. Microbiol. 152, 335-341, 1989

A:Title: Homology and distribution of CO dehydrogenase structural genes in carboxydotrop
A:Reference number: PL0138; MUID:90055678; PMID:2818128
A:Accession: PL0146
A:Molecule type: protein
A:Residues: 1-4 <KRA>

C:Comment: Carbon-monoxide dehydrogenase consists of three polypeptide chains: large, med
C:Keywords: oxidoreductase
Query Match 47.6%; Score 10; DB 2; Length 4;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 IAK 5

DB 1 MAK 3

RESULT 2

I40505
Hypothetical protein 3 (4 aa) - Bacillus stearothermophilus
C:Species: Bacillus stearothermophilus
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40505
R:Waye, M.M.; Winter, G.
Eur. J. Biochem. 158, 505-510, 1986
A:Title: A transcription terminator in the 5' non-coding region of the tyrosyl tRNA synt
A:Reference number: I40503; MUID:86274732; PMID:3525162
A:Accession: I40505
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4 <RES>
A:Cross-references: EMBL:X04193; NID:g40233; PIDN:CAA27783.1; PID:g580944

Query Match

Best Local Similarity 42.9%; Score 9; DB 2; Length 4;

Matches 1; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 VIAK 5

DB 1 MLSK 4

RESULT 3

A27897

glucan 1,4-alpha-glucosidase (EC 3.2.1.1.3) - Aspergillus phoenicis (fragment)
 N;Alternate names: glucoamylase
 C;Species: Aspergillus phoenicis
 C;Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 06-Dec-1996
 C;Accession: A27897
 R;Inokuchi, N.; Takahashi, T.; Irie, M.
 J. Biochem. 90, 1055-1067, 1981

A;Title: Purification and characterization of a minor glucoamylase from Aspergillus saitoi
 A;Reference number: A27897; MUID:82075730; PMID:6796572
 A;Note: Aspergillus saitoi
 A;Accession: A27897

A;Molecule type: protein
 A;Residues: 1-4 <INO>
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 42.9%; Score 9; DB 2; Length 4;
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;

Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVI 3
 :||
 Db 1 AVI 3

RESULT 4

S11127

phosphoprotein, bone - chicken (fragment)

C;Species: Gallus gallus (chicken)

C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 04-Mar-2000

C;Accession: S11127; S11128

R;Mikuni-Takagaki, Y.; Glimcher, M.J.

Biochem. J. 268, 585-591, 1990

A;Title: Post-translational processing of chicken bone phosphoproteins. Identification of
 A;Reference number: S11127; MUID:90303246; PMID:2363696

A;Accession: S11127

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-5 <MIK1>

A;Accession: S11128

A;Status: preliminary

A;Molecule type: protein

A;Residues: 'X', 2-5 <MIK2>

C;Keywords: phosphoprotein

Query Match 42.9%; Score 9; DB 2; Length 5;
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;

Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
 :||
 Db 3 VSK 5

RESULT 5

S17255

ribosomal protein Yml1, mitochondrial, questionable - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

A;Variety: strain 07173

C;Date: 23-Apr-1993 #sequence_revision 14-Sep-1994 #text_change 09-May-1997

C;Accession: S17255

R;Grohmann, L.; Graack, H.R.; Kruff, V.; Choli, T.; Goldschmidt-Reisin, S.; Kitakawa, M.

FEBS Lett. 284, 51-56, 1991

A;Title: Extended N-terminal sequencing of proteins of the large ribosomal subunit from

A;Reference number: S17255; MUID:91285106; PMID:2060626

A;Accession: S17255

A;Molecule type: protein

A;Residues: 1-4 <GRO>

C;Comment: A coding region for this protein could not be identified in the genome of Sac

C;Genetics:

A;Genome: nuclear

C;Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 38.1%; Score 8; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 2.8e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
 :||
 Db 1 SV 2

RESULT 6

T30569

hypothetical protein - Emericella nidulans

C;Species: Emericella nidulans, Aspergillus nidulans

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000

A;Accession: T30569

R;Morrice, J.; MacKenzie, D.A.; Parr, A.J.; Archer, D.B.

Curr. Genet. 34, 379-385, 1998

A;Title: Isolation and characterisation of the acetyl-CoA carboxylase gene from Aspergill

A;Reference number: Z20869; MUID:99087906; PMID:9871120

A;Accession: T30569

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-4 <MOR>

A;Cross-references: EMBL:Y15996; NID:e1285512; PID:e1218041; PIDN:CAA75927.1

Query Match 38.1%; Score 8; DB 2; Length 4;
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
 :||
 Db 3 SV 4

RESULT 7

E44823

synaptosomal-associated protein SNAP-25 peptide 1 - rabbit (fragment)

N;Alternate names: superprotein peptide 1

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 15-Jun-1996

C;Accession: E44823

R;Loewy, A.; Liu, W.S.; Baitinger, C.; Willard, M.B.

J. Neurosci. 11, 3412-3421, 1991

A;Title: The major 35S-methionine-labeled rapidly transported protein (superprotein) is

A;Reference number: A44823; MUID:92044785; PMID:1941090

A;Accession: E44823

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-4 <LOE>

A;Experimental source: visual tissue

A;Note: sequence extracted from NCBI backbone (NCBIP:64247)

C;Keywords: membrane trafficking

Query Match 38.1%; Score 8; DB 2; Length 4;
 Best Local Similarity 25.0%; Pred. No. 2.8e+05;

Matches 1; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VIAK 5
 :||
 Db 1 IMEX 4

RESULT 8

A60521

glycogen phosphorylase (EC 2.4.1.1), muscle - mullet (Liza ramada) (fragment)

N;Alternate names: glycogen phosphorylase b

C;Species: Liza ramada

C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Aug-2003

C;Accession: A60521

R;Bonamusa, L.; Baanante, I.V.

Comp. Biochem. Physiol. B 95, 295-301, 1990

A;Title: Purification and characterization of glycogen phosphorylase B from skeletal mus

A;Reference number: A60521; MUID:90227907; PMID:2103669

A;Accession: A60521

A:Molecule type: protein
A:Residues: 1-5 <BON>

C:Superfamily: glucan phosphorylase

F:3/Binding site: phosphate (Ser) (covalent) (by phosphorylase b kinase) #status experim

Query Match 38.1%; Score 8; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 2.8e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SV 2
||
3 SV 4

RESULT 9

TL4908

hypothetical protein - parsley

C:Species: Petroselinum crispum (parsley)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C:Accession: TL4908

R:Kircher, S.; Ledger, S.; Hayashi, H.; Weisshaar, B.; Schafer, E.; Frohnmeyer, H.

Mol. Gen. Genet. 257, 595-605, 1998

A:Title: CRF4a, a novel plant bZIP protein of the CTRF family: comparative analysis of

A:Reference number: Z18261; MUID:98265918; PMID:9604882

C:Accession: TL4908

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-5 <KIR>

A:Cross-references: EMBL:Y10809; NID:g3336901; PIDN:CAA71767.1; PID:g3336902

A:Experimental source: Hamburger Schnitt

Query Match 38.1%; Score 8; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 2.8e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SV 2
||
3 SV 4

RESULT 10

I61883

protamine P1 - orangutan (fragment)

C:Species: Pongo pygmaeus (orangutan)

C>Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000

C:Accession: I61883

R:Queralt, R.; Oliva, R.

Gene 133, 197-204, 1993

A:Title: Identification of conserved potential regulatory sequences of the protamine-enc

A:Reference number: I37013; MUID:94040810; PMID:8224908

C:Accession: I61883

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-4 <RES>

A:Cross-references: EMBL:Z12146; NID:g38156; PIDN:CAA78130.1; PID:g4379372

Query Match 33.3%; Score 7; DB 2; Length 4;

Best Local Similarity 33.3%; Pred. No. 2.8e+05;

Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 IAK 5
||
1 MAR 3

RESULT 11

PT0551

T-cell receptor beta chain V-D-J region (126-1CG) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997

C:Accession: PT0551

R:Feeney, A.J.

J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A:Reference number: PT0509; MUID:91277601; PMID:1711558

C:Accession: PT0551

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-4 <FEE>

A:Experimental source: day 18 fetal thymus, strain BALB/c

C:Keywords: T-cell receptor

Query Match 33.3%; Score 7; DB 2; Length 4;

Best Local Similarity 50.0%; Pred. No. 2.8e+05;

Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SV 2
||
3 SI 4

RESULT 12

PT0697

T-cell receptor beta chain V-D-J region (135-1BF) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997

C:Accession: PT0697

R:Feeney, A.J.

J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A:Reference number: PT0509; MUID:91277601; PMID:1711558

C:Accession: PT0697

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-4 <FEE>

A:Experimental source: newborn thymus, strain BALB/c

C:Keywords: T-cell receptor

Query Match 33.3%; Score 7; DB 2; Length 4;

Best Local Similarity 50.0%; Pred. No. 2.8e+05;

Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SV 2
||
2 SI 3

RESULT 13

I37013

protamine P1 - Cercopithecus patas (fragment)

C:Species: Cercopithecus patas

C>Date: 19-Mar-1997 #sequence_revision 07-Nov-1997 #text_change 21-Jul-2000

C:Accession: I37013

R:Queralt, R.; Oliva, R.

Gene 133, 197-204, 1993

A:Title: Identification of conserved potential regulatory sequences of the protamine-enc

A:Reference number: I37013; MUID:94040810; PMID:8224908

C:Accession: I37013

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-4 <RES>

A:Cross-references: EMBL:Z12150; NID:g22814; PIDN:CAA78134.1; PID:g4377415

Query Match 33.3%; Score 7; DB 2; Length 4;

Best Local Similarity 33.3%; Pred. No. 2.8e+05;

Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 IAK 5
||
1 MAR 3

RESULT 14

I84439

protamine P1 - savannah baboon (fragment)

C;Species: Papio hamadryas doquiera (savannah baboon)
 C;Date: 19-Mar-1997 #sequence_revision 07-Nov-1997 #text_change 21-Jul-2000
 C;Accession: I84439
 R;Queralt, R.; Oliva, R.
 Gene 133, 197-204, 1993
 A;Title: Identification of conserved potential regulatory sequences of the protamine-enc
 A;Reference number: I37013; MUID:94040810; PMID:8224908
 A;Accession: I84439
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-4 <RES>
 A;Cross-references: EMBL:Z12147; NID:G38134; PIDN:CAA78131.1; PID:G4379349

Query Match 33.3%; Score 7; DB 2; Length 4;
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;
 Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
 Db 1 MAR 3

RESULT 15
 I39964
 ribosomal protein S4 - Bacillus circulans (fragment)
 C;Species: Bacillus circulans
 C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 19-Jul-1996
 C;Accession: I39964
 R;Grundy, F.J.; Henkin, T.M.
 J. Bacteriol. 174, 6763-6770, 1992
 A;Title: Characterization of the Bacillus subtilis rpsD regulatory target site.
 A;Reference number: I39963; MUID:93015735; PMID:1400226
 A;Accession: I39964
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-5 <RES>
 A;Cross-references: GB:M99041; NID:G143471
 C;Genetics:
 A;Gene: rpsD

Query Match 33.3%; Score 7; DB 2; Length 5;
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;
 Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
 Db 1 MAR 3

Search completed: August 12, 2004, 06:55:20
 Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:21:13 ; Search time 13 Seconds
(without alignments)
20.027 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 38

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	47.6	4	1 DCMS_PSECH	P19918 pseudomonas
2	8	38.1	4	1 RM01_YEAST	P36515 saccharomyc
3	6	28.6	5	1 UF01_MOUSE	P38639 mus musculu
4	5	23.8	3	1 GRWM_HUMAN	P01157 homo sapien
5	5	23.8	3	1 LUXE_VIBRI	P24272 vibrio fisc
6	5	23.8	4	1 FEKA_ANGEL	P58705 anthopleura
7	5	23.8	4	1 TUFT_HUMAN	P01858 homo sapien
8	5	23.8	5	1 BPOB_CITFR	P12997 citrobacter
9	5	23.8	5	1 BPP7_BOTIN	P30425 bothrops in
10	5	23.8	5	1 B104_LITRU	P82100 litoria rub
11	5	23.8	5	1 TRM3_ECOLI	P13973 escherichia
12	4	19.0	4	1 ACHI_HUMAN	P35904 achatina fu
13	4	19.0	4	1 ECHI_HUMAN	P02731 homo sapien
14	4	19.0	4	1 FYRI_ANGEL	P58706 anthopleura
15	4	19.0	4	1 OCPI_COTMI	P58649 octopus min
16	4	19.0	5	1 AL14_CARMA	P81817 carcinus ma
17	4	19.0	5	1 B103_LITRU	P82099 litoria rub
18	4	19.0	5	1 FARP_ARTTR	P41853 artioposthi
19	4	19.0	5	1 PSK_DAUCA	P58261 daucus caro
20	4	19.0	5	1 RE11_LITRU	P82070 litoria rub
21	4	19.0	5	1 RE21_LITRU	P82071 litoria rub
22	4	19.0	5	1 RE31_LITRU	P82072 litoria rub
23	4	19.0	5	1 RE32_LITRU	P82073 litoria rub
24	4	19.0	5	1 SUGA_ACHDO	P19991 acheta dome
25	4	19.0	5	1 TPIS_CANPA	P54714 canis fami
26	4	19.0	5	1 UC22_MAIZE	P80628 zea mays (m
27	4	19.0	5	1 UXA4_CHLTR	P38005 chlamydia t
28	2	9.5	4	1 FAR3_HIRME	P42562 hirudo medi
29	2	9.5	4	1 FAR4_HIRME	P42563 hirudo medi
30	2	9.5	4	1 FLRF_HIRME	P42561 hirudo medi
31	2	9.5	4	1 FLRN_ANGEL	P58707 anthopleura
32	2	9.5	4	1 FNRF_MACNI	P01162 macrocallis
33	2	9.5	5	1 PRCT_PERAM	P01373 periplaneta

34	1	4.8	3	1	THYL_PIG	P01151 sus scrofa
35	1	4.8	4	1	DCML_PSECH	P19916 pseudomonas
36	1	4.8	5	1	BIOA_CITFR	P13071 citrobacter
37	0	0.0	4	1	OCPI_COTMI	P58648 octopus min
38	0	0.0	5	1	FAP2_PARMA	P81864 pardachirus

ALIGNMENTS

RESULT 1
DCMS_PSECH
ID DCMS_PSECH STANDARD; PRT; 4 AA.
AC P19918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbon monoxide dehydrogenase small chain (EC 1.2.99.2) (CO
DE dehydrogenase subunit S) (CO-DH S) (Fragment).
GN CUTS.
OS Pseudomonas carboxydohydrogena.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae.
OX NCBI_TaxID=290;
RN [1]
RP SEQUENCE.
RX MEDLINE=90055678; PubMed=2818128;
RA Kraut M., Hugendieck I., Herwig S., Meyer O.;
RT "Homology and distribution of CO dehydrogenase structural genes in
RT carboxydohydrophic bacteria.";
RL Microbiol. 152:335-341(1989).
CC -!- FUNCTION: Catalyzes the oxidation of carbon monoxide to carbon
CC dioxide.
CC -!- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
CC acceptor.
CC -!- COFACTOR: Binds 2 2Fe-2S clusters.
CC -!- SUBUNIT: CONSISTS OF THREE POLYPEPTIDE CHAINS: LARGE, MEDIUM, AND
CC SMALL.
DR PIR: P01146; P01146.
KW Oxidoreductase; Metal-binding; Iron-sulfur; Iron; 2Fe-2S.
FT NON TER 4 4
SQ SEQUENCE 4 AA; 420 MW; 6DD33DD6F0000000 CRC64;

Query Match 47.6%; Score 10; DB 1; Length 4;
Best Local Similarity 66.7%; Pred No. 1.4e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
Db 1 MAK 3

RESULT 2
RM01_YEAST
ID RM01_YEAST STANDARD; PRT; 4 AA.
AC P36515;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Mitochondrial 60S ribosomal protein L1 (Yml1) (Fragment).
GN MRPL1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE.
RX MEDLINE=91285106; PubMed=2060626;
RA Grohmann L., Graack H.-R., Kruff V., Choli T., Goldschmidt-Reisin S.,
RA Kitakawa M.;
RT "Extended N-terminal sequencing of proteins of the large ribosomal
RT subunit from yeast mitochondria.";
RL FEBS Lett. 284:51-56(1991).

DR PIR; S17255; S17255.
DR SGD; L0002681; MRPL1.
KW Ribosomal protein; Mitochondrion.
FT NON_TER 4 4
SQ SEQUENCE 4 AA; 402 MW; 7771B2D5D0000000 CRC64;
Query Match 38.1%; Score 8; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SV 2
DB 1 SV 2

RESULT 3
UF01_MOUSE STANDARD; PRT; 5 AA.
AC P38639;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Unknown protein from 2D-page of fibroblasts (P19) (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=95009907; PubMed=7523108;
RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
RT "Separation and sequencing of familial and novel murine proteins
RT using preparative two-dimensional gel electrophoresis."
RL Electrophoresis 15:735-745(1994).
CC -1- MISCELLANEOUS: On the 2D-gel the determined pI of this unknown
CC protein is: 6.6, its MW is: 19 kDa.
FT NON_TER 5 5
SQ SEQUENCE 5 AA; 717 MW; 7364087043100000 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 5;
Best Local Similarity 33.3%; Pred. No. 1.4e+05;
Matches 1; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 IAK 5
DB 2 IGR 4

RESULT 4
GRWM_HUMAN STANDARD; PRT; 3 AA.
AC P01157;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE Growth-modulating peptide.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=77162369; PubMed=858356;
RA Schlesinger D.H., Pickart L., Thaler M.M.;
RT "Growth-modulating serum tripeptide is glycyl-histidyl-lysine."
RL Experientia 33:324-325(1977).
CC -1- MISCELLANEOUS: This serum tripeptide has been found to stimulate
CC growth of some cell types and to inhibit other types in vitro.
DR GO; GO:0001558; P:regulation of cell growth; NAS.
SQ SEQUENCE 3 AA; 340 MW; 6331E81000000000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 3;

Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 K 5
DB 3 K 3

RESULT 5
LUXE_VIBFI STANDARD; PRT; 3 AA.
AC P24272;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Long-chain-fatty-acid-luciferin-component ligase (EC 6.2.1.19) (Acyl-
DE protein synthetase) (Fragment).
GN LUXE.
OS Vibrio fischeri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91072226; PubMed=2254256;
RA Swartzman E., Kapoor S., Graham A.F., Meighen E.A.;
RT "A new Vibrio fischeri lux gene precedes a bidirectional termination
RT site for the lux operon."
RL J. Bacteriol. 172:6797-6802(1990).
CC -1- FUNCTION: ACYL-PROTEIN SYNTHETASE ACTIVATES TETRADECANOIC ACID.
CC IT IS A COMPONENT OF THE FATTY ACID REDUCTASE COMPLEX RESPONSIBLE
CC FOR CONVERTING TETRADECANOIC ACID TO THE ALDEHYDE WHICH SERVES AS
CC SUBSTRATE IN THE LUCIFERASE-CATALYZED REACTION.
CC -1- CATALYTIC ACTIVITY: ATP + an acid + protein = AMP + diphosphate +
CC an acyl-protein thioester
CC -1- PATHWAY: Bioluminescent fatty acid reduction system; second step.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M62812; -; NOT ANNOTATED_CDS.
KW Luminescence; Ligase.
FT NON_TER 1 1
SQ SEQUENCE 3 AA; 374 MW; 6AA330300000000000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 3;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 K 5
DB 2 K 2

RESULT 6
FFKA_ANTEL STANDARD; PRT; 4 AA.
AC P58705;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antho-Kaamide.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RP SEQUENCE.

```

RX MEDLINE=92028852; PubMed=1681803;
RA Nothacker H.-P., Rinehart K.L. Jr., Grimmlikhuizen C.J.P.;
RT "Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-Xamide), a
RL novel neuropeptide from sea anemones.";
RN Biochem. Biophys. Res. Commun. 179:1205-1211(1991).
RN [2]
RP FUNCTION.
RX MEDLINE=93391436; PubMed=8397415;
RA McFarlane I.D., Hudman D., Nothacker H.-P., Grimmlikhuizen C.J.P.;
RT "The expansion behaviour of sea anemones may be coordinated by two
RT inhibitory neuropeptides, Antho-Xamide and Antho-Riamide.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).
CC -!- FUNCTION: Inhibits spontaneous contractions in several muscle
CC groups. May be involved in the expansion phase of feeding
CC behaviour in sea anemones.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Neuron specific.
DR PIR; JQ1273; JQ1273.
KW Neuropeptide; Amidation.
FT MOD_RES 1 1 L-3-PHENYLLACTYL.
FT MOD_RES 4 4 AMIDATION.
FT MOD_RES 5 5 DDD339C9A000000000 CRC64;
SQ SEQUENCE 4 AA; 512 MW; 6DD339C9A000000000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 K 5
Db 3 K 3

RESULT 7
TUFT_HUMAN
ID TUFT_HUMAN STANDARD; PRT; 4 AA.
AC P01859;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phagocytosis-stimulating peptide (Tuftsin).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=72187087; PubMed=4112769;
RA Nishioka K., Constantopoulos A., Satoh P.S., Najjar V.A.;
RT "The characteristics, isolation and synthesis of the phagocytosis
RT stimulating peptide tuftsin.";
RL Biochem. Biophys. Res. Commun. 47:172-179(1972).
RN [2]
RP IMMUNOGLOBULIN CLASS.
RX MEDLINE=68091045; PubMed=4169272;
RA Fidalgo B.V., Najjar V.A.;
RT "The physiological role of the lymphoid system. VI. The stimulatory
RT effect of leucophilic gamma globulin (leucokinin) on the phagocytic
RT activity of human polymorphonuclear leucocyte.";
RL Biochemistry 6:3386-3392(1967).
CC -!- MISCELLANEOUS: An IgG (called leucokinin) binds reversibly to the
CC cell membrane of neutrophils in the blood. Leucokininase on the
CC membrane releases the active peptide tuftsin from the gamma chain.
CC Tuftsin is essential for maximum stimulation of the phagocytic
CC activity of neutrophils.
DR PIR; A02147; A02147.
DR MIM; 191150; -.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006909; P:phagocytosis; NAS.
SQ SEQUENCE 4 AA; 501 MW; 74176321C000000000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 K 5
Db 2 K 2

RESULT 8
BIOB_CITR
ID BIOB_CITR STANDARD; PRT; 5 AA.
AC P12997;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase) (Fragment).
GN BIOB.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
CX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89006280; PubMed=2971595;
RA Shiuan D., Campbell A.;
RT "Transcriptional regulation and gene arrangement of Escherichia coli,
RT Citrobacter freundii and Salmonella typhimurium biotin operons.";
RL Gene 67:203-211(1988).
CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -!- PATHWAY: Biotin biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the biotin and lipoic acid synthetases
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21922; -, NOT_ANNOTATED_CDS.
DR PIR; I40698; I40698.
KW Biotin biosynthesis; Iron-sulfur; Transferase.
FT NON_TER 5
SQ SEQUENCE 5 AA; 532 MW; 75A5B1EDD6F00000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 5;
Best Local Similarity 50.0%; Pred. No. 1.4e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IA 4
Db 1 MA 2

RESULT 9
BPP7_BOTIN
ID BPP7_BOTIN STANDARD; PRT; 5 AA.
AC P30425;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bradykinin-potentiating peptide S5,2 (5A) (Angiotensin-converting
DE enzyme inhibitor).
OS Bothrops insularis (Island jararaca) (Queimada jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
CX NCBI_TaxID=8723;
RN [1]
RP SEQUENCE.
RX TISSUE=Venom;
RX MEDLINE=90351557; PubMed=2386615;
RA Cintra A.C.O., Vieira C.A., Giglio J.R.;

```

RT "Primary structure and biological activity of bradykinin potentiating
 RL J. Protein Chem. 9:221-227(1990).
 CC -1- FUNCTION: This peptide both inhibits the activity of the
 CC angiotensin-converting enzyme and enhances the action of
 CC bradykinin by inhibiting the kinases that inactivate it.
 CC It acts as an indirect hypotensive agent.
 DR PIR; G37196; G37196.
 KW Hypotensive agent; Pyrrolidone carboxylic acid.
 FT MOD RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 SQ SEQUENCE 5 AA; 629 MW; 776DC37326B00000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred.No. 1.4e+05;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 K 5
 Db 2 K 2

RESULT 10

ID_E104_LITRU STANDARD; PRT; 5 AA.
 AC P82100;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Electrin 4.
 OS Litoria rubella (Desert tree frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 CC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae;
 CC Pelodyadinae; Litoria.
 OX NCBI_TaxID=104895;
 RN [1]
 RP SEQUENCE.

RC TISSUE=Skin secretion;
 RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
 RT "Peptides from the skin glands of the Australian buzzing tree frog
 RT Litoria electrica. Comparison with the skin peptides from Litoria
 RT rubella".
 RL Aust. J. Chem. 52:639-645(1999).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Skin.
 KW Amphibian defense peptide; Amidation.
 FT MOD RES 5 5 AMIDATION.
 SQ SEQUENCE 5 AA; 616 MW; 61F2D1A059A00000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 5;
 Best Local Similarity 50.0%; Pred.No. 1.4e+05;
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
 Db 3 IV 4

RESULT 11

TRM3_ECOLI STANDARD; PRT; 5 AA.
 ID_TRM3_ECOLI
 AC P13973;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE TrAM protein (Fragment).
 GN TRAM.
 OS Escherichia coli.
 OG Plasmid IncFII R100.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=88227859; PubMed=2836369;
 RA Inamoto S., Yoshioka Y., Ohtsubo E.;
 RT "Identification and characterization of the products from the traJ
 RT and traY genes of plasmid R100".
 RL J. Bacteriol. 170:2749-2757(1988).
 CC -1- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION
 CC PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the traM family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M20941; -; NOT_ANNOTATED_CDS.
 DR PIR; A32014; A32014.
 KW Conjugation; Plasmid; DNA-binding.
 FT NON TER 1 1
 SQ SEQUENCE 5 AA; 634 MW; 6B1B1AA443500000 CRC64;

Query Match 23.8%; Score 5; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred.No. 1.4e+05;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 K 5
 Db 1 K 1

RESULT 12

ACH1_ACHFU STANDARD; PRT; 4 AA.
 ID_ACH1_ACHFU
 AC P35904;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Achatin-I.
 OS Achatina fulica (Giant African snail).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
 CC Sigmurethra; Achatinoidea; Achatinidae; Achatina.
 OX NCBI_TaxID=6530;
 RN [1]
 RP SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.
 RC STRAIN=Ferussac; TISSUE=Ganglion;
 RX MEDLINE=89273551; PubMed=2597281;
 RA Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,
 RA Funase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,
 RA Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;
 RT "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina
 RT fulica Ferussac containing a D-amino acid residue".
 RL Biochem. Biophys. Res. Commun. 160:1015-1020(1989).
 RN [2]
 RP CHARACTERIZATION.

RC STRAIN=Ferussac; TISSUE=Heart atrium;
 RX MEDLINE=91264856; PubMed=1675568;
 RA Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,
 RA Yoshida M., Harada A., Muneoka Y., Kobayashi M.;
 RT "Purification of achatin-I from the atria of the African giant snail,
 RT Achatina fulica, and its possible function".
 RL Biochem. Biophys. Res. Commun. 177:847-853(1991).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=93014529; PubMed=1399265;
 RA Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,
 RA Iwashita T., Nomoto K.;
 RT "Crystal structure and molecular conformation of achatin-I
 RT (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a
 RT D-amino acid residue".
 RL Int. J. Pept. Protein Res. 39:258-264(1992).

CC -!- FUNCTION: Neuroexcitatory peptide; increases the impulse frequency
 CC and produces a spike broadening of the identified heart excitatory
 CC neuron (PON); also enhances the amplitude and frequency of the
 CC heart beat. Has also an effect on several other muscles.

DR PIR; A32480; A32480.

KW Hormone; D-amino acid.

FT MOD RES 2 2 D-PHENYLALANINE.

SQ SEQUENCE 4 AA; 408 MW; 6AADD9C810000000 CRC64;

Query Match 19.0%; Score 4; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 1.4e+05;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 A 4

Db 3 A 3

RESULT 13

EOSI HUMAN

ID _EOSI HUMAN STANDARD; PRT; 4 AA.

AC P02731;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 21-JUL-1986 (Rel. 01, Last annotation update)

DE Eosinophilotoxic peptides.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE.

RX MEDLINE=76078412; PubMed=1060093;

RA Goetzl E.J., Austen K.F.;

RT "Purification and synthesis of eosinophilotoxic tetrapeptides of
 RT human lung tissue: identification as eosinophil chemotactic factor of
 RT anaphylaxis.";

RL Proc. Natl. Acad. Sci. U.S.A. 72:4123-4127(1975).

CC -!- MISCELLANEOUS: These peptides are released from mast cells in lung
 CC (and other tissues) during hypersensitivity reactions

CC (anaphylaxis). Their activities, preferentially affecting

CC eosinophils, include chemotaxis, chemotactic deactivation, release

CC of enzymes, and stimulation of the hexose monophosphate shunt.

CC GO; GO:0006935; P:chemotaxis; IDA.

DR GO; GO:0006955; P:immune response; IDA.

FT VARIANT 1 1 V -> A (IN OTHER PEPTIDE).

FT /FTID=VAR_005201.

SQ SEQUENCE 4 AA; 390 MW; 6B05B862A0000000 CRC64;

Query Match

Best Local Similarity 100.0%; Score 4; DB 1; Length 4;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 V 2

Db 1 V 1

RESULT 14

FYRI ANTEL

ID _FYRI ANTEL STANDARD; PRT; 4 AA.

AC P58706;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Antho-Riamide I [Contains: Antho-Riamide II].

OS Anthopleura elegantissima (Sea anemone).

OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;

OC Nynanthaeae; Actiniidae; Anthopleura.

OX NCBI_TaxID=6110;

RN [1]

RP SEQUENCE.

RX MEDLINE=92270459; PubMed=1821096;

RA Nothacker H.-P., Rinehart K.L. Jr., McFarlane I.D.,
 RA Grimmelikhuizen C.J.P.;
 RT "Isolation of two novel neuropeptides from sea anemones: the unusual,
 RT biologically active L-3-phenylalanyl-Tyr-Arg-Ile-NH2 and its
 RT des-phenylalanyl fragment Tyr-Arg-Ile-NH2.";
 RL Peptides 12:1165-1173(1991).

RN [2]

RP FUNCTION.

RX MEDLINE=93391436; PubMed=8397415;

RA McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuizen C.J.P.;

RT "The expansion behaviour of sea anemones may be coordinated by two

RT inhibitory neuropeptides, Antho-Riamide and Antho-Riamide.";

RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).

CC -!- FUNCTION: Inhibits spontaneous contractions in several muscle

CC groups. May be involved in the expansion phase of feeding

CC behaviour in sea anemones.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: Neuron specific.

KW Neuropeptide; Amidation.

FT CHAIN 1 4 ANTHO-RIAMIDE I.

FT CHAIN 2 4 ANTHO-RIAMIDE II.

FT MOD RES 1 1 L-3-PHENYLALANYL.

FT MOD RES 4 4 AMIDATION.

SQ SEQUENCE 4 AA; 598 MW; 60441B59A0000000 CRC64;

Query Match

Best Local Similarity 100.0%; Score 4; DB 1; Length 4;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 I 3

Db 4 I 4

RESULT 15

OCF3 OCTMI

ID _OCF3 OCTMI STANDARD; PRT; 4 AA.

AC P58649;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cardioactive peptides Ocp-3/Ocp-4.

OS Octopus minor (Octopus).

OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;

OC Octopodiformes; Octopoda; Incirrata; Octopodidae; Octopus.

OX NCBI_TaxID=89766;

RN [1]

RP SEQUENCE, SYNTHESIS, MASS SPECTROMETRY, AND CHARACTERIZATION.

RT TISSUE=Brain;

RX MEDLINE=20336815; PubMed=10876044;

RA Wakoshi E., Hisada M., Minakata H.;

RT "Cardioactive peptides isolated from the brain of a Japanese octopus,

RT Octopus minor.";

RL Peptides 21:623-630(2000).

CC -!- FUNCTION: Cardioactive; has both positive chronotropic and

CC inotropic effects on the heart. Ocp-4 is a 1000 time less

CC active than Ocp-3.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- PTM: Ocp-4 has D-Ser instead of L-Ser.

CC -!- MASS SPECTROMETRY: MW=395.2; METHOD=MALDI.

KW Hormone; D-amino acid.

FT MOD RES 2 2 D-SERINE (IN OCP-4).

SQ SEQUENCE 4 AA; 463 MW; 6AB365B810000000 CRC64;

Query Match

Best Local Similarity 100.0%; Score 4; DB 1; Length 4;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 S 1

Db 2 S 2

Search completed: August 12, 2004, 06:53:27
Job time : 14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:32:33 ; Search time 75 Seconds
(without alignments)
21.035 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp archea: *
2: sp bacteria: *
3: sp fungi: *
4: sp human: *
5: sp invertebrate: *
6: sp mammal: *
7: sp mhc: *
8: sp organelle: *
9: sp phage: *
10: sp plant: *
11: sp rodent: *
12: sp virus: *
13: sp vertebrate: *
14: sp unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	33.3	4	11	Q08433 rattus sp.
2	5	23.8	4	5	P83568 sepia offic
3	5	23.8	5	2	P83073 bacillus ce
4	5	23.8	5	10	Q99007 hordeum vul
5	2	9.5	5	13	P83308 gallus gall
6	0	0.0	2	5	P83570 sepia offic

ALIGNMENTS

RESULT 1
Q08433
ID Q08433
AC Q08433
DT 01-NOV-1996 (Tremblrel. 01, Created)

PRELIMINARY; PRT; 4 AA.

01-NOV-1996 (Tremblrel. 01, Last sequence update)
01-JUN-2003 (Tremblrel. 24, Last annotation update)
Bilirubin UDP-glucuronosyltransferase (Fragment).
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10118;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Gunn;
RX MEDLINE=91282758; PubMed=1840486;
RA Sato H., Aono S., Kashiwamata S., Koizumi O.;
RT "Genetic defect of bilirubin UDP-glucuronosyltransferase in the
hyperbilirubinemic Gunn rat.";
RL Biochem. Biophys. Res. Commun. 177:1161-1164(1991).
DR EMBL; S38636; AAB19259.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
KW Transferase.
FT NON TER 1 1
SQ SEQUENCE 4 AA; 473 MW; 633732C420000000 CRC64;

Query Match 33.3%; Score 7; DB 11; Length 4;
Best Local Similarity 33.3%; Pred.No. 1e+06;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVI 3
Db 1 NVL 3

RESULT 2

P83568
ID P83568
AC P83568
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Pheromone peptide ILME.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OX NCBI_TaxID=6610;
[1]
RN SEQUENCE, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND MASS
RP SPECTROMETRY.
RC TISSUE=Egg;
RX PubMed=10944467;
RA Zatylny C., Gagnon J., Boucaud-Camou E., Henry J.;
RT "ILME: a waterborne pheromonal peptide released by the eggs of Sepia
officinalis.";
RL Biochem. Biophys. Res. Commun. 275:217-222(2000).
[2]
RN SEQUENCE.
RP TISSUE=Egg;
RX PubMed=12207899;
RA Zatylny C., Marvin L., Gagnon J., Henry J.;
RT "Fertilization in Sepia officinalis: the first mollusk sperm-
attracting peptide.";
RL Biochem. Biophys. Res. Commun. 296:1186-1193(2002).
CC -!- FUNCTION: HAS MYOTROPIC ACTIVITY TARGETING THE GENITAL TRACT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: FOLLICLE, FULLY GROWN OOCYTE AND EGG(EC2).
CC -!- MASS SPECTROMETRY: MW=505.4; METHOD=MALDI.
DR GO; GO:0005186; F:pheromone activity; IEA.
KW Pheromone.
SQ SEQUENCE 4 AA; 505 MW; 6B16972030000000 CRC64;

Query Match 23.8%; Score 5; DB 5; Length 4;
Best Local Similarity 0.0%; Pred.No. 1e+06;
Matches 0; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VI 3
Db 1 NVL 3

```
Db 1 IL 2
RESULT 3
P83073 ID P83073 PRELIMINARY; PRT; 5 AA.
AC P83073;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 88 kDa protein (fragment).
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE.
RC STRAIN=NCIMB 11796;
RA Browne N., Dows B.C.A.;
RL Submitted (JUL-2001) to Swiss-Prot.
FT NON TER 5
SQ SEQUENCE 5 AA; 623 MW; 6B01AAA336F00000 CRC64;

Query Match 23.8%; Score 5; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 K 5
Db 2 K 2

RESULT 4
Q99007 ID Q99007 PRELIMINARY; PRT; 5 AA.
AC Q99007;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha amylase (Fragment).
GN AMY1 GENE.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91329704; PubMed=1831055;
RA Jacobsen J.V., Close T.J.;
RT "Control of transient expression of chimaeric genes by gibberellic
RT acid and abscisic acid in protoplasts prepared from mature barley
RT aleurone layers.";
RL Plant Mol. Biol. 16:713-721(1991).
DR EMBL; X54643; CAA38455.1; -.
FT NON_TER 5
SQ SEQUENCE 5 AA; 600 MW; 61E3344DD6F00000 CRC64;

Query Match 23.8%; Score 5; DB 10; Length 5;
Best Local Similarity 50.0%; Pred. No. 1e+06;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IA 4
Db 1 MA 2

RESULT 5
P83308 ID P83308 PRELIMINARY; PRT; 5 AA.
AC P83308;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE FMRamide-like neuropeptide (LPLRF-amide).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE=Brain;
RX PubMed=6137771;
RA Dockray G.J., Reeve J.R. Jr., Shively J., Gayton R.J., Barnard C.S.;
RT "A novel active pentapeptide from chicken brain identified by
RT antibodies to FMRamide.";
RL Nature 305:328-330(1983).
CC -!- FUNCTION: MAY FUNCTION AS A NEUROTRANSMITTER OR MODULATOR.
CC -!- SIMILARITY: BELONGS TO THE FARP (FMRAMIDE RELATED PEPTIDE)
CC FAMILY.
DR GO; GO:0007218; P:neuropeptide signaling pathway; TAS.
KW Neuropeptide; Amidation.
FT MOD RES 5
SQ SEQUENCE 5 AA; 645 MW; 69D4073767400000 CRC64;

Query Match 9.5%; Score 2; DB 13; Length 5;
Best Local Similarity 0.0%; Pred. No. 1e+06;
Matches 0; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 I 3
Db 1 L 1

RESULT 6
P83570 ID P83570 PRELIMINARY; PRT; 2 AA.
AC P83570;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Neuropeptide GWA.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OX NCBI_TaxID=6610;
RN [1]
RP SEQUENCE, FUNCTION, MASS SPECTROMETRY, AND AMIDATION.
RC TISSUE=Optic lobe;
RX PubMed=9437704;
RA Henry J., Favrel P., Boucaud-Camou E.;
RT "Isolation and identification of a novel Ala-Pro-Gly-Trp-amide-related
RT peptide inhibiting the motility of the mature oviduct in the
RT cuttlefish, Sepia officinalis.";
RL Peptides 18:1469-1474(1997).
CC -!- FUNCTION: REGULATORY NEUROPEPTIDE WITH MYOTROPIC ACTIVITY
CC TARGETING THE DISTAL OVIDUCT. INHIBITS THE MOTILITY OF THE OVIDUCT
CC BY DECREASING TONUS, FREQUENCY AND AMPLITUDE OF CONTRACTIONS.
CC -!- MASS SPECTROMETRY: MW=259.9; METHOD=MALDI.
DR GO; GO:0007218; P:neuropeptide signaling pathway; IEA.
KW Neuropeptide; Amidation.
FT MOD RES 2
SQ SEQUENCE 2 AA; 261 MW; 73781000000000000000 CRC64;

Query Match 0.0%; Score 0; DB 5; Length 2;
Best Local Similarity 0.0%; Pred. No. 1e+06;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 S 1
Db 1 G 1

Search completed: August 12, 2004, 06:54:52
Job time : 75 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:20:13 ; Search time 90 Seconds
(without alignments)
15.697 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 34717

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	21	100.0	5	3	AAY97147 Pigment p
2	21	100.0	5	5	Abb99061 N-termina
3	16	76.2	5	2	Aar72928 Yeast PPI
4	16	76.2	5	2	Aar72927 Yeast PPI
5	16	76.2	5	2	Aar72884 Yeast PPI
6	16	76.2	5	6	Aao29981 Peptide #
7	16	76.2	5	7	Abr63446 Rat purin
8	15	71.4	5	1	Aap97806 Sequence
9	14	66.7	4	2	Aar61324 Fragment
10	14	66.7	5	2	Aay07986 Human sec
11	13	61.9	4	1	Aap91629 Motif use
12	13	61.9	4	1	Aap97808 Sequence
13	13	61.9	4	2	Aaw55770 Immunisa
14	13	61.9	4	5	Abb84333 Human MBP
15	13	61.9	4	7	Abr57300 Thermus o
16	13	61.9	5	2	Aar12661 Pentapept
17	13	61.9	5	2	Aar51525 Mimotope
18	13	61.9	5	2	Aar69893 Pentameri
19	13	61.9	5	2	Aar78989 p107 pept
20	13	61.9	5	2	Aar98639 Peptide 1
21	12	57.1	4	2	Aar15757 Farnesyl
22	12	57.1	4	2	Aar49753 Farnesyl
23	12	57.1	4	2	Aar77816 Farnesyl
24	12	57.1	4	2	Aaw04443 Farnesyl
25	12	57.1	4	2	Aaw65412 Peptide u

26	12	57.1	4	2	AAY28344	Peptide f
27	12	57.1	4	3	AAY87947	Mammalian
28	12	57.1	4	4	AAG65468	Substrate
29	12	57.1	4	4	AAB57512	Mannose r
30	12	57.1	4	4	AAB57922	Mannose r
31	12	57.1	4	4	AAB55665	Monocyte
32	12	57.1	4	4	AAB80566	Peptide u
33	12	57.1	4	5	ABJ05144	T-cell su
34	12	57.1	4	5	ABP63437	Monocyte
35	12	57.1	4	5	AAE20561	Soybean d
36	12	57.1	4	6	ABU79151	Prenylati
37	12	57.1	4	7	ADC26827	Anti-angi
38	12	57.1	4	7	ADD11758	T cell su
39	12	57.1	5	1	AAp61368	N-termina
40	12	57.1	5	2	AAR11930	Pentapept
41	12	57.1	5	2	AAR51510	Mimotope
42	12	57.1	5	2	AAR71699	pBSmutlac
43	12	57.1	5	2	AAR71698	pBSmutlac
44	12	57.1	5	2	AAR69878	Pentameri
45	12	57.1	5	2	AAR66898	Agonist p

ALIGNMENTS

RESULT 1
AAY97147
ID AAY97147 standard; peptide; 5 AA.
XX AC AAY97147;
XX AC AAY97147;
DT 04-DEC-2000 (first entry)
XX AC AAY97147;
DE Pigment protein from coral tissue N-terminal peptide 1.
XX AC AAY97147;
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX AC AAY97147;
OS Acropora aspera.
OS Acropora horrida.
OS Montipora caliculata.
OS Montipora monasteriata.
OS Porites murrayensis.
OS Porites lobata.
XX AC AAY97147;
PN WO200046233-A1.
XX AC AAY97147;
PD 10-AUG-2000.
XX AC AAY97147;
PF 02-FEB-2000; 2000WO-AU0000056.
XX AC AAY97147;
PR 02-FEB-1999; 99AU-00008463.
XX AC AAY97147;
PA (UNSY) UNIV SYDNEY.
XX AC AAY97147;
PI Hoegh-Guldberg O, Dove S;
XX AC AAY97147;
DR WPI; 2000-532892/48.
XX AC AAY97147;
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX AC AAY97147;
PS Claim 3; Page 42; 49pp; English.
XX AC AAY97147;
CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters

CC (both claimed)
 XX Sequence 5 AA;
 SQ Sequence 5 AA;
 Query Match 100.0%; Score 21; DB 3; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SVIAK 5
 Db 1 SVIAK 5
 RESULT 2
 ID ABB99061 standard; peptide; 5 AA.
 AC ABB99061;
 XX 22-JAN-2003 (first entry)
 DT N-terminal amino acid sequence of a CFM #1.
 DE Colour facilitating molecule; CFM; green fluorescent protein; GFP;
 XX chromophore; biomatrix; transgenic animal; colouring agent;
 KW flower industry; expression marker; reporter molecule; photon trap;
 KW UV sink; sunscreen.
 XX Unidentified.
 OS WO200270703-A2.
 PN 12-SEP-2002.
 XX 01-MAR-2002; 2002WO-GB000928.
 XX 02-MAR-2001; 2001US-0273227P.
 PR 21-MAR-2001; 2001AU-00003874.
 PR 15-OCT-2001; 2001US-0329816P.
 XX (NUFA-) NUFARM LTD.
 PA (UYQU) UNIV QUEENSLAND.
 PA (JONE)/ JONES E L.
 XX Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
 PI Hoegh-Guldberg IO, Prescott M;
 XX WPI: 2002-740765/80.
 DR Novel color-facilitating molecule for producing a biomatrix, has a
 PT polypeptide which alone/along with molecules imparts altered visual
 PT characteristics to cells in the absence of excitation by extraneous non-
 PT white light.
 XX Claim 3; Page 278; 510pp; English.
 FS The invention relates to an isolated colour-facilitating molecule (CFM)
 CC comprising a polypeptide which, in a cell, alone or together with one or
 CC more other molecules imparts an altered visual characteristic to the cell
 CC when visualised by a human eye in the absence of excitation by extraneous
 CC non-white light or particle emission. CFMs are useful for producing a
 CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
 CC red coloured fleece. They are useful for producing coloured plant
 CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
 CC uses include transducing or intensifying an image, providing additional
 CC light for growing phototropic organisms e.g. algae and/or corals, for
 CC coating materials that experience UV damage e.g. plastics and car
 CC upholstery. CFMs are useful in the flower industry, in the development of
 CC new varieties of flowering plants. Other contemplated uses include,
 CC expression markers, general reporter molecules, photon traps, UV sinks or
 CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
 CC fungal species, and in fruits and vegetables to enhance their
 CC marketability. CFMs embedded in a gel matrix improve image quality in

CC situations of distorted light spectra (biomatrix). The first all-protein
 CC chromophore to be isolated was Green Fluorescent protein (GFP). The
 CC current sequence represents the N-terminal amino acid sequence of a
 CC colour-facilitating molecule (CFM)
 XX Sequence 5 AA;
 SQ Sequence 5 AA;
 Query Match 100.0%; Score 21; DB 5; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SVIAK 5
 Db 1 SVIAK 5
 RESULT 3
 AAR72928
 ID AAR72928 standard; peptide; 5 AA.
 XX AAR72928;
 AC AAR72928;
 XX 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 29-NOV-1995 (first entry)
 XX Yeast PPIase tryptic fragment 8.
 DE Escherichia coli; protein conformation; folding; acceleration;
 XX PPIase-alpha; peptidyl prolyl cis trans isomerase alpha; catalysis;
 KW isomerisation; prolyl peptide bond.
 XX Saccharomyces cerevisiae; strain AH22.
 XX EP647714-A1.
 PN 12-APR-1995.
 PD 19-JUL-1990; 94EP-00203612.
 PF 19-JUL-1989; 89JP-00184738.
 PR 06-OCT-1989; 89JP-00260244.
 PR 29-DEC-1989; 89JP-00344705.
 PR 19-JUL-1990; 90EP-00307914.
 XX (TOFU) TONEN CORP.
 PA Hayano T, Katou S, Maki N, Takahashi N, Suzuki M;
 XX WPI: 1995-140756/19.
 DR New E.coli peptidyl prolyl cis trans isomerase beta - used to accelerate
 XX the folding of proteins, partic. for activation of inactive recombinant
 XX proteins.
 XX Example 2; Page 23; 85pp; English.
 FS AAR72921-29 are tryptic fragments of a yeast pPIase (peptidyl prolyl cis
 CC trans isomerase). The yeast pPIase has a single mol. wt. of about 17 kDa
 CC and a single isoelectric point of about 6.2. The enzyme catalyses the
 CC isomerisation of prolyl peptide bonds in proteins and accelerates the
 CC folding of the protein. The inventors are claiming a pPIase-beta.
 CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
 CC correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
 CC on 16-OCT-2003 to standardise OS field)
 XX Sequence 5 AA;
 SQ Sequence 5 AA;
 Query Match 76.2%; Score 16; DB 2; Length 5;
 Best Local Similarity 75.0%; Pred. No. 1.4e+06;
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 VIAK 5

```

Db      |:||
        2 VVAK 5

RESULT 4
AAR72927
ID AAR72927 standard; peptide; 5 AA.
XX
AC AAR72927;
XX
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 29-NOV-1995 (first entry)
XX
DE Yeast PPIase tryptic fragment 7.
XX
KW Escherichia coli; protein conformation; folding; acceleration;
KW PPIase-alpha; peptidyl prolyl cis trans isomerase alpha; catalysis;
KW isomerisation; prolyl peptide bond.
XX
OS Saccharomyces cerevisiae; strain AH22.
XX
PN EP647714-A1.
XX
PD 12-APR-1995.
XX
PF 19-JUL-1990; 94EP-00203612.
XX
PR 19-JUL-1989; 89JP-00184738.
PR 06-OCT-1989; 89JP-00260244.
PR 29-DEC-1989; 89JP-00344705.
PR 19-JUL-1990; 90EP-00307914.
XX
PA (TOFU ) TONEN CORP.
XX
PI Hayano T, Katou S, Maki N, Takahashi N, Suzuki M;
XX
PD WPI; 1995-140756/19.
XX
PF New E.coli peptidyl prolyl cis trans isomerase alpha - used to accelerate
PT the folding of proteins, partic. for activation of inactive recombinant
PT proteins.
XX
PS Example 2; Page 23; 85pp; English.
XX
CC AAR72877-85 are tryptic fragments of a yeast PPIase (peptidyl prolyl cis
CC trans isomerase). The yeast PPIase has a single mol. wt. of about 17 kDa
CC and a single isoelectric point of about 6.2. The enzyme catalyses the
CC isomerisation of prolyl peptide bonds in proteins and accelerates the
CC folding of the protein. The inventors are claiming a PPIase-beta.
CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 5 AA;

Query Match 76.2%; Score 16; DB 2; Length 5;
Best Local Similarity 75.0%; Pred. No. 1.4e+06;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 VVAK 5
        |:||
Db      2 VVAK 5

RESULT 6
AAO29981
ID AAO29981 standard; peptide; 5 AA.
XX
AC AAO29981;
XX
DT 03-SEP-2003 (first entry)
XX
DE Peptide #4 used to enhance cellular adherence and growth.
XX
KW Cellular adhesion; growth; expression; secretion.
XX
OS Unidentified.
XX
PN WO200304045-A2.
XX
PD 30-MAY-2003.
XX
PF 19-NOV-2002; 2002WO-US037207.
XX

```


ID AAR61324 standard; peptide; 4 AA.
 AC AAR61324;
 XX
 DT 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 27-APR-1995 (first entry)
 XX
 XX Fragment of deacetylase enzyme.
 DE
 XX
 XX Deacetylase; enzyme; L-N-acetylphosphinothricin; L-AcPPT;
 KW L-phosphinothricin; PPT; glutamine synthase; plant; male sterility;
 KW anther.
 XX
 XX Variovorax paradoxus; (mixed culture).
 OS Brevundimonas diminuta; (mixed culture).
 OS Nocardia globerula; (mixed culture).
 OS Cellulosimicrobium cellulans; (mixed culture).
 OS Agrobacterium tumefaciens; (mixed culture).
 XX
 XX DE4308061-A1.
 PN
 XX
 XX 15-SEP-1994.
 ED
 XX
 XX 13-MAR-1993; 93DE-04308061.
 PF
 XX
 XX 13-MAR-1993; 93DE-04308061.
 PR
 XX
 XX (FARH) HOECHST AG.
 PA
 XX
 XX Schulz A, Bartsch K;
 PI
 XX
 XX WPI; 1994-2866683/36.
 DR
 XX
 XX New de-acetylase specific for L-N-acetyl-phosphinothricin - isolated from
 PT soil microbes, useful for stereoselective prodn. of L-phosphinothricin
 PT and for making male-sterile plants.
 PT
 XX
 XX Claim 5; Page 5; 5pp; German.
 PS
 XX
 XX The deacetylase has a molecular weight of 20000-100000, an optimum pH of
 CC 6.5-9.5 and substrate specificity for L-N-acetylphosphinothricin (L-
 CC AcPPT). It may be used for the deacetylation of AcPPT for the
 CC stereoselective production of L-phosphinothricin (PPT) and for inducing
 CC reversible male sterility in plants (PPT inhibits the enzyme glutamine
 CC synthase in anthers). The deacetylase comprises at least one of four
 CC sequences (See AAR61321-24). (Updated on 25-MAR-2003 to correct PN
 CC field.) (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 4 AA;
 Query Match 66.7%; Score 14; DB 2; Length 4;
 Best Local Similarity 75.0%; Pred. No. 1.4e+06;
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 VIAK 5
 Db 1 VMAK 4
 RESULT 10
 ID AAY07986
 ID AAY07986 standard; protein; 5 AA.
 XX
 AC AAY07986;
 XX
 XX 06-JUL-1999 (first entry)
 DT
 XX
 XX Human secreted protein fragment #3 encoded from gene 41.
 DE
 XX
 KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;

KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 XX arthritis; malignancy; digestive; endocrine; infection.
 OS Homo sapiens.
 XX
 PN WO9918208-A1.
 XX
 XX 15-APR-1999.
 PD
 XX
 XX 01-OCT-1998; 98WO-US020775.
 PF
 XX
 XX 02-OCT-1997; 97US-0060833P.
 PR 02-OCT-1997; 97US-0060836P.
 PR 02-OCT-1997; 97US-0060837P.
 PR 02-OCT-1997; 97US-0060838P.
 PR 02-OCT-1997; 97US-0060839P.
 PR 02-OCT-1997; 97US-0060843P.
 PR 02-OCT-1997; 97US-0060862P.
 PR 02-OCT-1997; 97US-0060866P.
 PR 02-OCT-1997; 97US-0060874P.
 PR 02-OCT-1997; 97US-0060880P.
 PR 02-OCT-1997; 97US-0060884P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Duan DR, Florence KA, Ruben SM, Greene JM, Young P;
 PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
 PI Lafleur DW, Shi Y;
 XX
 XX WPI; 1999-264022/22.
 DR
 XX
 XX New isolated human genes and the secreted polypeptides they encode.
 PT
 XX
 XX Disclosure; Page 344; 368pp; English.
 PS
 XX
 XX This invention describes novel isolated human genes and the secreted
 CC proteins they encode. The products of the invention are useful for
 CC preventing, treating or ameliorating medical conditions, e.g. by protein
 CC or gene therapy. Also pathological conditions can be diagnosed by
 CC determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 101 polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours,
 CC neurodegenerative disorders, developmental abnormalities and fetal
 CC deficiencies, blood disorders, leukemias, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
 CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
 CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
 CC transplant rejection, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
 CC infections and AIDS. The human secreted proteins of the invention are
 CC represented in AAY07852-Y07993 and the encoding nucleic acids are
 CC represented in AAX37451-X37552
 XX
 SQ Sequence 5 AA;
 Query Match 66.7%; Score 14; DB 2; Length 5;
 Best Local Similarity 40.0%; Pred. No. 1.4e+06;
 Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVIAK 5
 Db 1 SMVSK 5
 RESULT 11
 ID AAP91629
 ID AAP91629 standard; protein; 4 AA.
 XX

CC host. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-
CC 2003 to correct PR field.)
XX Sequence 4 AA;
SQ

Query Match 61.9%; Score 13; DB 2; Length 4;
Best Local Similarity 50.0%; Pred. No. 1.4e+06;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIAK 5
Db 1 LVAK 4

RESULT 14
ABB84333
ID ABB84333 standard; peptide; 4 AA.
XX ABB84333;
AC
XX
DT 17-OCT-2002 (first entry)
DE Human MBP protein derived peptide SEQ ID 33.
XX
KW MBP; myelin basic protein; human; tolerance; immune system;
KW multiple sclerosis; autoimmune response; autoimmune disease;
KW immunosuppressive; neuroprotective.
XX
OS Homo sapiens.
XX
FN US2002076412-A1.
XX
PD 20-JUN-2002.
XX
PF 07-JUN-1995; 95US-00484409.
XX
PR 17-AUG-1987; 87US-00086694.
PR 12-JUL-1989; 89US-00379500.
PR 01-MAY-1990; 90US-00517245.
PR 01-MAY-1991; 91WO-US002991.
PR 30-APR-1992; 92US-00877444.
PR 21-MAY-1993; 93US-00066325.
PR 22-SEP-1993; 93US-00125407.
XX
PA (STEI/) STEINMAN L.
PA (ZAMV/) ZAMVIL S.
XX
FI Steinman L, Zamvil S;
XX
DR - WPI; 2002-598709/64.
XX
PT Modulating or tolerizing the immune system, useful for treating multiple
PT sclerosis, by administering a peptide derived from human myelin binding
PT protein.
XX
PS Disclosure; Page 14; 21pp; English.
XX
CC This invention describes a novel method for modulating or tolerizing the
CC immune system, and for treating multiple sclerosis comprising
CC administering a peptide derived from hMBP (human myelin basic protein).
CC The peptide induces an autoimmune response (T cell) to a self-antigen (or
CC part of it), and binds to an MHC (major histocompatibility complex)
CC antigen of a host susceptible to autoimmune diseases, i.e. competes with
CC binding to MBP and inhibit proliferation of MBP-reactive cells. The
CC peptide has immunosuppressive and neuroprotective activity. This sequence
CC represents a peptide derived from the human MBP protein which can be used
CC for tolerization
XX
SQ Sequence 4 AA;

Query Match 61.9%; Score 13; DB 5; Length 4;
Best Local Similarity 50.0%; Pred. No. 1.4e+06;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIAK 5
Db 1 LVAK 4

RESULT 15
ABR57300
ID ABR57300 standard; peptide; 4 AA.
XX
AC ABR57300;
XX
DT 09-SEP-2003 (first entry)
DE Thermus oshimai nucleic acid polymerase peptide 704-707 SEQ ID NO:26.
XX
KW Thermus oshimai; nucleic acid polymerase; enzyme; DNA sequencing;
KW amplification; reverse transcription; RNA amplification;
KW primer extension.
XX
OS Thermus oshimai.
XX
FN WO2003048310-A2.
XX
PD 12-JUN-2003.
XX
PF 22-NOV-2002; 2002WO-US037764.
XX
PR 30-NOV-2001; 2001US-0334798P.
XX
PA (APPL-) APPLERA CORP.
XX
FI Bolchakova E, Rozzelle J;
XX
DR WPI; 2003-505286/47.
XX
PT New nucleic acid, useful for DNA sequencing or amplification, reverse
PT transcription, RNA amplification or primer extension reactions.
XX
PS Disclosure; Page 32; 64pp; English.
XX
CC The present invention describes an isolated nucleic acid (I) encoding a
CC nucleic acid polymerase or a derivative nucleic acid polymerase with a
CC mutation that decreases 5-3' exonuclease activity or that reduces
CC discrimination against dideoxynucleotide triphosphates. Also described:
CC (1) a vector comprising the nucleic acid (I); (2) a host cell comprising
CC the nucleic acid (I); (3) a nucleic acid polymerase or its derivative;
CC (4) a kit comprising a container containing the nucleic acid polymerase
CC of (3); (5) making the nucleic acid polymerase of (3); (6) synthesizing a
CC DNA; (7) thermocyclic amplification of nucleic acid; and (8) primer
CC extending a DNA. The nucleic acid (I) is useful for DNA sequencing or
CC amplification, reverse transcription, RNA amplification or primer
CC extension reactions. The present sequence represents a Thermus oshimai
CC nucleic acid polymerase peptide, which is given in the exemplification of
CC the present invention
XX
SQ Sequence 4 AA;

Query Match 61.9%; Score 13; DB 7; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IAK 5
Db 2 IAK 4

Search completed: August 12, 2004, 06:53:04
Job time : 93 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:54:59 ; Search time 41 Seconds
(without alignments)
38,284 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 17582

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	76.2	5	10	US-09-992-124A-4
2	14	66.7	5	10	US-09-862-145A-4
3	14	66.7	5	10	US-09-862-145A-8
4	14	66.7	5	10	US-09-862-145A-12
5	14	66.7	5	10	US-09-862-145A-16
6	14	66.7	5	10	US-09-862-145A-20
7	14	66.7	5	10	US-09-862-145A-24
8	14	66.7	5	15	US-10-195-730-329
9	14	66.7	5	15	US-10-402-029-4
10	14	66.7	5	15	US-10-402-029-8
11	14	66.7	5	15	US-10-402-029-12
12	14	66.7	5	15	US-10-402-029-16
13	14	66.7	5	15	US-10-402-029-20
14	14	66.7	5	15	US-10-402-029-24
15	14	66.7	5	16	US-10-285-108A-4

16	14	66.7	5	16	US-10-285-108A-8	Sequence 8, Appli
17	14	66.7	5	16	US-10-285-108A-12	Sequence 12, Appli
18	14	66.7	5	16	US-10-285-108A-16	Sequence 16, Appli
19	14	66.7	5	16	US-10-285-108A-20	Sequence 20, Appli
20	14	66.7	5	16	US-10-285-108A-24	Sequence 24, Appli
21	13	61.9	4	8	US-08-484-409-33	Sequence 33, Appli
22	13	61.9	4	14	US-10-303-109A-26	Sequence 26, Appli
23	13	61.9	5	14	US-10-197-927-17	Sequence 17, Appli
24	12	57.1	4	9	US-09-797-543-3	Sequence 3, Appli
25	12	57.1	4	9	US-09-359-325A-31	Sequence 31, Appli
26	12	57.1	4	9	US-09-945-249-38	Sequence 38, Appli
27	12	57.1	4	9	US-09-870-759-163	Sequence 163, App
28	12	57.1	4	10	US-09-751-708A-163	Sequence 163, App
29	12	57.1	4	12	US-10-371-406B-3	Sequence 3, Appli
30	12	57.1	4	13	US-10-016-717-4	Sequence 4, Appli
31	12	57.1	4	13	US-10-033-026-11	Sequence 11, Appli
32	12	57.1	4	13	US-10-078-458-8	Sequence 8, Appli
33	12	57.1	4	14	US-10-087-905-6	Sequence 6, Appli
34	12	57.1	4	14	US-10-287-639-2	Sequence 2, Appli
35	12	57.1	4	14	US-10-087-942-6	Sequence 6, Appli
36	12	57.1	4	14	US-10-083-894-18	Sequence 18, Appli
37	12	57.1	4	14	US-10-083-894-42	Sequence 42, Appli
38	12	57.1	5	8	US-08-859-699-20	Sequence 20, Appli
39	12	57.1	5	9	US-09-764-246-5	Sequence 5, Appli
40	12	57.1	5	9	US-09-875-519A-29	Sequence 29, Appli
41	12	57.1	5	9	US-09-359-325A-30	Sequence 30, Appli
42	12	57.1	5	9	US-09-748-114-28	Sequence 28, Appli
43	12	57.1	5	9	US-09-945-249-37	Sequence 37, Appli
44	12	57.1	5	10	US-09-992-124A-3	Sequence 3, Appli
45	12	57.1	5	10	US-09-992-124A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-992-124A-4
; Sequence 4, Application US/09992124A
; Publication No. US20030162289A1
; GENERAL INFORMATION:
; APPLICANT: Heidaran, Mohammad A.
; APPLICANT: Haaland, Perry D.
; APPLICANT: Wilkins, Jamie H.
; APPLICANT: Spargo, Catherine A.
; APPLICANT: Campbell, Robert L.
; TITLE OF INVENTION: Peptides Promoting Cell Adherence, Growth and Secretion
; FILE REFERENCE: 102-410
; CURRENT APPLICATION NUMBER: US/09/992,124A
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide selected for biological activity
US-09-992-124A-4

Query Match 76.2%; Score 16; DB 10; Length 5;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 VIAX 5
|:|
Db 2 VWAX 5

RESULT 2

US-09-862-145A-4
; Sequence 4, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:

```

; APPLICANT: Seiberg, Miri
; APPLICANT: Shapiro, Stanley
; TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
; FILE REFERENCE: J&J-1991
; CURRENT APPLICATION NUMBER: US/09/862,145A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: MISC FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: C-terminal Amidation
US-09-862-145A-4

Query Match          66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SLICK 5

RESULT 3
US-09-862-145A-8
; Sequence 8, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
; APPLICANT: Seiberg, Miri
; APPLICANT: Shapiro, Stanley
; TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
; FILE REFERENCE: J&J-1991
; CURRENT APPLICATION NUMBER: US/09/862,145A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-09-862-145A-8

Query Match          66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SLICK 5

RESULT 4
US-09-862-145A-12
; Sequence 12, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
; APPLICANT: Seiberg, Miri
; APPLICANT: Shapiro, Stanley
; TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
; FILE REFERENCE: J&J-1991
; CURRENT APPLICATION NUMBER: US/09/862,145A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-09-862-145A-12

Query Match          66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SLICK 5

RESULT 5
US-09-862-145A-16
; Sequence 16, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
; APPLICANT: Seiberg, Miri
; APPLICANT: Shapiro, Stanley
; TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
; FILE REFERENCE: J&J-1991
; CURRENT APPLICATION NUMBER: US/09/862,145A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Palmitoyl N-terminus
US-09-862-145A-16

Query Match          66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SLICK 5

RESULT 6
US-09-862-145A-20
; Sequence 20, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
; APPLICANT: Seiberg, Miri
; APPLICANT: Shapiro, Stanley
; TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
; FILE REFERENCE: J&J-1991
; CURRENT APPLICATION NUMBER: US/09/862,145A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-09-862-145A-20

Query Match          66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
Db 1 SLICK 5

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Stearatoyl N-terminus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: Stearatoyl C-terminus
US-09-862-145A-20

Query Match 66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 7

US-09-862-145A-24
; Sequence 24, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
; APPLICANT: Seiberg, Miri
; APPLICANT: Shapiro, Stanley
; TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
; FILE REFERENCE: J&J-1991
; CURRENT APPLICATION NUMBER: US/09/862,145A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Stearatoyl N-terminus
US-09-862-145A-24

Query Match 66.7%; Score 14; DB 10; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 8

US-10-195-730-329
; Sequence 329, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02

; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 329
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-195-730-329

Query Match 66.7%; Score 14; DB 14; Length 5;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIAK 5
|:|
Db 1 SMVSK 5

RESULT 9

US-10-402-029-4
; Sequence 4, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Seiberg, Miri
; TITLE OF INVENTION: Compositions for Darkening the Skin
; FILE REFERENCE: J&J-2099
; CURRENT APPLICATION NUMBER: US/10/402,029
; CURRENT FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: C-terminal Amidation
US-10-402-029-4

Query Match 66.7%; Score 14; DB 15; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 10

US-10-402-029-8
; Sequence 8, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Seiberg, Miri
; TITLE OF INVENTION: Compositions for Darkening the Skin
; FILE REFERENCE: J&J-2099
; CURRENT APPLICATION NUMBER: US/10/402,029
; CURRENT FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-402-029-8

Query Match 66.7%; Score 14; DB 15; Length 5;

Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 11

US-10-402-029-12
; Sequence 12, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Seiberg, Miri
; TITLE OF INVENTION: Compositions for Darkening the Skin
; FILE REFERENCE: J&J-2099
; CURRENT APPLICATION NUMBER: US/10/402,029
; CURRENT FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Palmitoyl N-terminus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: Amidated C-terminus
US-10-402-029-12

Query Match 66.7%; Score 14; DB 15; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 12

US-10-402-029-16
; Sequence 16, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Seiberg, Miri
; TITLE OF INVENTION: Compositions for Darkening the Skin
; FILE REFERENCE: J&J-2099
; CURRENT APPLICATION NUMBER: US/10/402,029
; CURRENT FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Palmitoyl N-terminus
US-10-402-029-16

Query Match 66.7%; Score 14; DB 15; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;

Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 13

US-10-402-029-20
; Sequence 20, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Seiberg, Miri
; TITLE OF INVENTION: Compositions for Darkening the Skin
; FILE REFERENCE: J&J-2099
; CURRENT APPLICATION NUMBER: US/10/402,029
; CURRENT FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Stearatoyl N-terminus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: Amidated C-terminus
US-10-402-029-20

Query Match 66.7%; Score 14; DB 15; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 14

US-10-402-029-24
; Sequence 24, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Seiberg, Miri
; TITLE OF INVENTION: Compositions for Darkening the Skin
; FILE REFERENCE: J&J-2099
; CURRENT APPLICATION NUMBER: US/10/402,029
; CURRENT FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Stearatoyl N-terminus
US-10-402-029-24

Query Match 66.7%; Score 14; DB 15; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SVIAK 5
|:|
Db 1 SLIGK 5

RESULT 15
US-10-285-108A-4
; Sequence 4, Application US/10285108A
; Publication No. US20040091449A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Connie
; APPLICANT: Wu, Jane
; TITLE OF INVENTION: Compositions for Darkening the Skin and/or Hair
; FILE REFERENCE: J&J-2172
; CURRENT APPLICATION NUMBER: US/10/285,108A
; CURRENT FILING DATE: 2002-10-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (5)..(5)
; OTHER INFORMATION: C-terminal Amidation
US-10-285-108A-4

Query Match 66.7%; Score 14; DB 16; Length 5;
Best Local Similarity 60.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SVIAK 5
|:|
Db 1 SLIGK 5

Search completed: August 12, 2004, 07:02:19
Job time : 41 secs

This page blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:51:29 ; Search time 18 Seconds
(without alignments)
14,341 Million cell updates/sec

Title: US-09-890-463-1

Perfect score: 21

Sequence: 1 SVIAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 24558

Minimum DB seq length: 0

Maximum DB seq length: 5

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	61.9	5	1	US-08-193-977-16
2	13	61.9	5	6	Sequence 16, Appl Patent No. 5217869
3	12	57.1	4	1	US-07-969-305-66
4	12	57.1	4	1	Sequence 66, Appl
5	12	57.1	4	1	US-07-969-305-67
6	12	57.1	4	1	Sequence 67, Appl
7	12	57.1	4	1	US-08-456-424-124
8	12	57.1	4	2	Sequence 124, Appl
9	12	57.1	4	2	Sequence 25, Appl
10	12	57.1	4	3	Sequence 3, Appl
11	12	57.1	4	3	Sequence 19, Appl
12	12	57.1	4	4	US-08-842-306B-19
13	12	57.1	4	4	Sequence 3, Appl
14	12	57.1	4	4	US-08-838-973B-17
15	12	57.1	4	4	Sequence 17, Appl
16	12	57.1	4	4	US-09-294-987-4
17	12	57.1	4	4	Sequence 4, Appl
18	12	57.1	4	4	US-09-268-163-11
19	12	57.1	4	4	Sequence 11, Appl
20	12	57.1	4	4	Sequence 8, Appl
21	12	57.1	4	4	US-09-601-178-2
22	12	57.1	4	4	Sequence 2, Appl
23	12	57.1	4	4	US-09-797-543-3
24	12	57.1	4	4	Sequence 3, Appl
25	12	57.1	4	4	US-09-665-362A-18
26	12	57.1	4	4	Sequence 42, Appl
27	12	57.1	4	5	PCT-US93-08062-25
28	12	57.1	5	1	US-08-363-475-13
29	12	57.1	5	1	Sequence 25, Appl
30	12	57.1	5	1	US-07-969-305-50
31	12	57.1	5	1	Sequence 50, Appl
32	12	57.1	5	1	US-07-969-305-51
33	12	57.1	5	1	Sequence 51, Appl
34	12	57.1	5	1	US-07-969-305-52
35	12	57.1	5	1	Sequence 52, Appl
36	12	57.1	5	1	US-08-097-938-49
37	12	57.1	5	1	Sequence 49, Appl
38	12	57.1	5	1	US-08-476-000-49
39	12	57.1	5	1	Sequence 49, Appl
40	12	57.1	5	2	US-08-472-840-49
41	12	57.1	5	2	US-08-441-871-138
42	12	57.1	5	2	Sequence 138, Appl
43	12	57.1	5	2	US-08-476-976-49
44	12	57.1	5	3	US-08-474-410-49
45	12	57.1	5	3	Sequence 49, Appl

28	12	57.1	5	3	US-09-061-768A-5	Sequence 5, Appl
29	12	57.1	5	3	US-09-142-334-29	Sequence 29, Appl
30	12	57.1	5	3	US-08-842-306B-18	Sequence 18, Appl
31	12	57.1	5	3	US-08-838-973B-16	Sequence 16, Appl
32	12	57.1	5	3	US-08-486-673B-49	Sequence 49, Appl
33	12	57.1	5	4	US-08-964-747-4	Sequence 4, Appl
34	12	57.1	5	4	US-09-562-913-4	Sequence 4, Appl
35	12	57.1	5	4	US-09-764-246-5	Sequence 5, Appl
36	12	57.1	5	6	5217869-4	Patent No. 5217869
37	11	52.4	4	1	US-08-014-979-18	Sequence 18, Appl
38	11	52.4	4	1	US-08-014-979-25	Sequence 25, Appl
39	11	52.4	4	1	US-08-142-439A-8	Sequence 8, Appl
40	11	52.4	4	2	US-08-869-477-8	Sequence 8, Appl
41	11	52.4	4	2	US-08-691-997-12	Sequence 12, Appl
42	11	52.4	4	3	US-08-859-242-18	Sequence 18, Appl
43	11	52.4	4	3	US-09-171-554-10	Sequence 10, Appl
44	11	52.4	4	3	US-09-171-554-11	Sequence 11, Appl
45	11	52.4	4	4	US-08-492-411A-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-193-977-16
; Sequence 16, Application US/08193977
; Patent No. 5625031
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, KEVIN R.
; APPLICANT: COLEMAN, KEVIN G.
; TITLE OF INVENTION: PEPTIDE INHIBITORS OF THE P33CDK2 AND
; TITLE OF INVENTION: P34CDC2 CELL CYCLE REGULATORY KINASES AND HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS E7 ONCOPROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,977
; FILING DATE: 08-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-193-977-16

Query Match 61.9%; Score 13; DB 1; Length 5;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIAK 5

Db 2 IMAK 5

RESULT 2
 5217869-19
 ; PATENT NO. 5217869
 ; APPLICANT: KAUFAR, LAWRENCE M.
 ; TITLE OF INVENTION: METHOD TO PRODUCE IMMUNODIAGNOSTIC
 ; REAGENTS
 ; NUMBER OF SEQUENCES: 121
 ; CURRENT APPLICATION DATA: US/07/255.906
 ; FILING DATE: 11-OCT-1988
 ; SEQ ID NO:19:
 ; LENGTH: 5
 5217869-19

Query Match 61.9%; Score 13; DB 6; Length 5;
 Best Local Similarity 75.0%; Pred. No. 3e+05;
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVIA 4
 Db 2 SMIA 5

RESULT 3
 US-07-969-305-66
 ; Sequence 66, Application US/07969305
 ; Patent No. 5609872
 ; GENERAL INFORMATION:
 ; APPLICANT: AHLBORG, Niklas
 ; APPLICANT: BERZINS, Klavs
 ; APPLICANT: PERLMANN, Peter
 ; TITLE OF INVENTION: NEW PEPTIDES AND THEIR USE
 ; NUMBER OF SEQUENCES: 70
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis
 ; STREET: George Mason Bldg., Washington & Prince Sts.
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/969.305
 ; FILING DATE: 08-APR-1993
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: SE 9002684-0
 ; FILING DATE: 17-AUG-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Crane-Feury, Sharon E
 ; STREET: George Mason Bldg., Washington & Prince Sts.
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/969.305
 ; FILING DATE: 08-APR-1993
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: SE 9002684-0
 ; FILING DATE: 17-AUG-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Crane-Feury, Sharon E
 ; REGISTRATION NUMBER: 36,113
 ; REFERENCE/DOCKET NUMBER: 003300-286
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 836-6620
 ; TELEFAX: (703) 836-2021
 ; INFORMATION FOR SEQ ID NO: 66:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-07-969-305-66

Query Match 57.1%; Score 12; DB 1; Length 4;
 Best Local Similarity 100.0%; Pred. No. 3e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVI 3
 Db 1 SVI 3

RESULT 5
 US-08-456-424-124
 ; Sequence 124, Application US/08456424
 ; Patent No. 5807979
 ; GENERAL INFORMATION:
 ; APPLICANT: SATTERTHWAIT JR., ARNOLD C.
 ; APPLICANT: ARRHENIUS, THOMAS
 ; APPLICANT: CABEZAS, EDELMIRA
 ; TITLE OF INVENTION: SYNTHETIC, STABILIZED, THREE-DIMENSION
 ; TITLE OF INVENTION: POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 145
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER

Query Match 57.1%; Score 12; DB 1; Length 4;
 Best Local Similarity 100.0%; Pred. No. 3e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVI 3
 Db 2 SVI 4

RESULT 4
 US-07-969-305-67
 ; Sequence 67, Application US/07969305
 ; Patent No. 5609872
 ; GENERAL INFORMATION:
 ; APPLICANT: AHLBORG, Niklas
 ; APPLICANT: BERZINS, Klavs
 ; APPLICANT: PERLMANN, Peter
 ; TITLE OF INVENTION: NEW PEPTIDES AND THEIR USE
 ; NUMBER OF SEQUENCES: 70
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis
 ; STREET: George Mason Bldg., Washington & Prince Sts.
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/969.305
 ; FILING DATE: 08-APR-1993
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: SE 9002684-0
 ; FILING DATE: 17-AUG-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Crane-Feury, Sharon E
 ; REGISTRATION NUMBER: 36,113
 ; REFERENCE/DOCKET NUMBER: 003300-286
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 836-6620
 ; TELEFAX: (703) 836-2021
 ; INFORMATION FOR SEQ ID NO: 67:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-07-969-305-67

Query Match 57.1%; Score 12; DB 1; Length 4;
 Best Local Similarity 100.0%; Pred. No. 3e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVI 3
 Db 1 SVI 3

RESULT 5
 US-08-456-424-124
 ; Sequence 124, Application US/08456424
 ; Patent No. 5807979
 ; GENERAL INFORMATION:
 ; APPLICANT: SATTERTHWAIT JR., ARNOLD C.
 ; APPLICANT: ARRHENIUS, THOMAS
 ; APPLICANT: CABEZAS, EDELMIRA
 ; TITLE OF INVENTION: SYNTHETIC, STABILIZED, THREE-DIMENSION
 ; TITLE OF INVENTION: POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 145
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER

```

; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,424
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/224,059
; FILING DATE:
; APPLICATION NUMBER: US 07/866,040
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BOZICEVIC, KARL
; REGISTRATION NUMBER: 28,807
; REFERENCE/DOCKET NUMBER: 278022000120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 124:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-456-424-124

Query Match 57.1%; Score 12; DB 1; Length 4;
Best Local Similarity 25.0%; Pred. No. 3e+05;
Matches 1; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIA 5
Db 1 IVSK 4

RESULT 6
US-08-429-964-25
; Sequence 25, Application US/08429964
; Patent No. 5962243
; GENERAL INFORMATION:
; APPLICANT: BROWN, MICHAEL S.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: REISS, YUVAL
; APPLICANT: JAMES, GUY L.
; TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF FARNESYL
; TITLE OF INVENTION: TRANSFERASE INHIBITORS
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,964
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/021,625
; FILING DATE: 16-FEB-1993
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/822,011
; FILING DATE: ABANDONED
; CLASSIFICATION: 435
; APPLICATION NUMBER: PCT/US/91/02650
; FILING DATE: 18-APR-1991
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/615,715
; FILING DATE: 20-NOV-1990
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/510,706
; FILING DATE: 18-APR-1990 (ABANDONED)
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PARKER, DAVID L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:432/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-429-964-25

Query Match 57.1%; Score 12; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIA 4
Db 2 VIA 4

RESULT 7
US-09-173-887-3
; Sequence 3, Application US/09173887
; Patent No. 6245884
; GENERAL INFORMATION:
; APPLICANT: Hook, Vivian Y.H.
; TITLE OF INVENTION: SECTASES RELATED TO ALZHEIMER'S DEMENTIA
; FILE REFERENCE: P-AS 3337
; CURRENT APPLICATION NUMBER: US/09/173,887
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: mammalian
US-09-173-887-3

Query Match 57.1%; Score 12; DB 3; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIA 4
Db 1 VIA 3

RESULT 8
US-08-842-306B-19
; Sequence 19, Application US/08842306B
; Patent No. 6271197
; GENERAL INFORMATION:
; APPLICANT: Berlin, Vivian

```

```

Levin, David
Ohya, Yoshikazu
Damagnez, Veronique
Smith, Susan
TITLE OF INVENTION: ASSAYS AND REAGENTS FOR IDENTIFYING
ANTI-FUNGAL AGENTS, AND USES RELATED THERETO
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordpad
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/842.306B
FILING DATE: 23-Apr-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/771,212
FILING DATE: 20-DEC-1996
APPLICATION NUMBER: US 08/631,319
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV-074.04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-7000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-842-306B-19
Query Match 57.1%; Score 12; DB 3; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps
QY 2 VIA 4
Db 2 VIA 4
RESULT 9
US-08-838-973B-17
Sequence 17, Application US/08838973B
Patent No. 6277564
GENERAL INFORMATION:
APPLICANT: Berlin, Vivian
Damagnez, Veronique
Smith, Susan
TITLE OF INVENTION: ASSAYS AND REAGENTS FOR IDENTIFYING
ANTI-FUNGAL AGENTS, AND USES RELATED THERETO
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

; CURRENT APPLICATION NUMBER: US/09/268,163B
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 60/077,901
; EARLIER FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Conus geographus
US-09-268-163-11

Query Match 57.1%; Score 12; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIA 4
|||
Db 2 VIA 4

RESULT 12
US-09-308-175A-8
; Sequence 8, Application US/09308175A
; Patent No. 6355617
; GENERAL INFORMATION:
; APPLICANT: LUKE, Richard William Arthur
; APPLICANT: COTTON, Ronald
; TITLE OF INVENTION: PEPTIDE DERIVATIVES
; FILE REFERENCE: 1991-174
; CURRENT APPLICATION NUMBER: US/09/308,175A
; CURRENT FILING DATE: 1999-05-17
; PRIOR APPLICATION NUMBER: PCT/GB97/03199
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: GB 9624562.6
; PRIOR FILING DATE: 1996-11-27
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-308-175A-8

Query Match 57.1%; Score 12; DB 4; Length 4;
Best Local Similarity 66.7%; Pred. No. 3e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IAK 5
:||
Db 1 VAK 3

RESULT 13
US-09-601-178-2
; Sequence 2, Application US/09601178
; Patent No. 6548306
; GENERAL INFORMATION:
; APPLICANT: ADMON, Arie
; APPLICANT: PALTIELI, Yoav
; APPLICANT: MANDELI, Silvia
; APPLICANT: SLOTKY, Ronit
; TITLE OF INVENTION: PLACENTAL PROTEIN 13
; FILE REFERENCE: ADMON=1
; CURRENT APPLICATION NUMBER: US/09/601,178
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: PCT/IL99/00036
; PRIOR FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: IL 123098
; PRIOR FILING DATE: 1998-01-29

; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-601-178-2

Query Match 57.1%; Score 12; DB 4; Length 4;
Best Local Similarity 75.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VIAK 5
|||
Db 1 VIIK 4

RESULT 14
US-09-797-543-3
; Sequence 3, Application US/09797543
; Patent No. 6627409
; GENERAL INFORMATION:
; APPLICANT: Hook, Vivian Y.H.
; TITLE OF INVENTION: SECRETASES RELATED TO ALZHEIMER'S DEMENTIA
; FILE REFERENCE: P-AS 4579
; CURRENT APPLICATION NUMBER: US/09/797,543
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 09/173,887
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: mammalian
US-09-797-543-3

Query Match 57.1%; Score 12; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIA 4
|||
Db 1 VIA 3

RESULT 15
US-09-665-362A-18
; Sequence 18, Application US/09665362A
; Patent No. 6632626
; GENERAL INFORMATION:
; APPLICANT: BROWN, MICHAEL S.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: REISS, YUVAL
; TITLE OF INVENTION: METHODS OF ASSAYING FARNESYL TRANSFERASE
; FILE REFERENCE: UTSD:249USD1
; CURRENT APPLICATION NUMBER: US/09/665,362A
; CURRENT FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 07/937,893
; PRIOR FILING DATE: 1992-12-22
; PRIOR APPLICATION NUMBER: 07/615,715
; PRIOR FILING DATE: 1990-11-20
; PRIOR APPLICATION NUMBER: 07/510,706
; PRIOR FILING DATE: 1990-04-18
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide

US-09-665-362A-18

Query Match 57.1%; Score 12; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VIA 4
|
|
|
Db 2 VIA 4

Search completed: August 12, 2004, 06:55:49
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 07:01:35 ; Search time 37 seconds
(without alignments)
44.196 Million cell updates/sec

Title: US-09-890-463-2

Perfect score: 83

Sequence: 1 SVIAKQMTYKVMGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 2990

Minimum DB seq length: 0

Maximum DB seq length: 17

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	31.3	16	2 B60278	24K antigen - Myco
2	22	26.5	11	2 C49037	Tcr gamma V-J regi
3	22	26.5	14	2 A47421	leukotriene B-4 12
4	22	26.5	15	2 PH1616	Ig H chain V-D-J r
5	22	26.5	16	2 B45895	T-cell surface gly
6	22	26.5	17	2 S78756	ribosomal protein
7	21	25.3	12	2 C49215	urease (EC 3.5.1.5
8	21	25.3	16	2 S09732	photosystem I prot
9	21	25.3	17	2 B42965	talain (glycosylate
10	21	25.3	17	2 S6164	ferredoxin al - Ja
11	20	24.1	8	2 S63493	dissimilatory sulf
12	20	24.1	9	2 G41946	T-cell receptor ga
13	20	24.1	9	2 G85802	hypothetical prote
14	20	24.1	10	2 S66458	ferredoxin - Rhizo
15	20	24.1	12	2 H41946	T-cell receptor ga
16	20	24.1	13	2 PA0049	protein QAL00046 -
17	20	24.1	14	2 PH1617	Ig H chain V-D-J r
18	20	24.1	15	2 S03353	plastocyanin - Mic
19	20	24.1	15	2 P50251	15K protein 5106 -
20	19	22.9	11	2 B49037	Tcr gamma V-J regi
21	19	22.9	12	2 PN0162	malate dehydrogena
22	19	22.9	12	2 A37933	Ig lambda chain J
23	19	22.9	13	2 PN0125	serine proteinase
24	19	22.9	13	2 PS0325	tetrahydroberlin
25	19	22.9	14	2 A60158	disaggregatase - M
26	19	22.9	14	2 PA0101	protein Q200020 -
27	19	22.9	15	2 S21241	oligo-1,6-glucosid
28	19	22.9	15	2 S21240	alpha-glucosidase
29	19	22.9	15	2 S21202	glucan 1,4-alpha-g

30	19	22.9	15	2 PA0099	phenotypic variati
31	19	22.9	15	2 A27504	histone H2A - mous
32	19	22.9	16	2 F41299	T-cell receptor al
33	19	22.9	16	2 S24667	protein-tyrosine k
34	19	22.9	16	2 S55307	glutathione transf
35	19	22.9	17	2 A27486	foliitropin inhibi
36	18	21.7	9	2 I46016	cytokeraatin 4 - bo
37	18	21.7	11	2 PN0167	ribosomal protein
38	18	21.7	11	2 PN0214	T-cell receptor be
39	18	21.7	12	2 I60434	68kDa neurofilamen
40	18	21.7	12	2 PH1611	Ig H chain V-D-J r
41	18	21.7	13	2 J02309	hypothetical 1.6K
42	18	21.7	13	2 J02319	hypothetical 1.6K
43	18	21.7	13	2 B20907	Ig kappa-1 chain J
44	18	21.7	15	2 A41338	isocitrate lyase (
45	18	21.7	15	2 S43321	RNA-binding protei

ALIGNMENTS

RESULT 1

B60278

24K antigen - Mycobacterium bovis (fragment)

C:Species: Mycobacterium bovis

C:Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 18-Jun-1993

C:Accession: B60278

R:Efifis, T.; Costopoulos, C.; Radford, A.J.; Bacic, A.; Wood, P.R.

Infect. Immun. 59, 800-807, 1991

A:Title: Purification and characterization of major antigens from a Mycobacterium bovis c

A:Reference number: A60278; MUID:91147217; PMID:1900061

A:Accession: B60278

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-16 <P>

Query Match

Best Local Similarity 31.3%; Score 26; DB 2; Length 16;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 TYKVMGSGT 16

||| : ||

Db 4 TYKEELKGT 12

RESULT 2

C49037

Tcr gamma V-J region - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: C49037

R:Ezquerria, A.; Wilde, D.B.; McConnell, T.J.; Sturmhofel, K.; Valas, R.B.; Shevach, E.M.;

Eur. J. Immunol. 22, 491-498, 1992

A:Title: Mouse autoreactive gamma/delta T cells. II. Molecular characterization of the T

A:Reference number: A49037; MUID:92164730; PMID:1311262

A:Accession: C49037

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-11 <EQ>

A:Cross-references: GB:S90639; NID:G246292; PIDN:AAB21549.1; PID:G246293

A:Experimental source: dendritic epidermal T-cell lines

A:Note: sequence extracted from NCBI backbone (NCBIN:90639, NCBI:P:90645)

Query Match

Best Local Similarity 26.5%; Score 22; DB 2; Length 11;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 11 TYMMSGT 16

|:| |

Db 3 VVMGGT 8

RESULT 3

Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 YKVMYS 14
:|:|:
Db 4 FKTYLS 9

RESULT 9

B42965
talin (glycosylated sites) - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: B42965
R:Hagmann, J.; Grob, M.; Burger, M.M.
J. Biol. Chem. 267, 14424-14428, 1992
A:Title: The cytoskeletal protein talin is O-glycosylated.
A:Reference number: A42965; PMID:92332560; PMID:1629228
A:Accession: B42965
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-17 <HAG>
A:Experimental source: gizzard
A>Note: sequence extracted from NCBI backbone (NCBIP:108592)

Query Match 25.3%; Score 21; DB 2; Length 17;
Best Local Similarity 42.9%; Pred. No. 2.9e+03;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 VIAKQMT 8
:|:|:
Db 2 ILANQLT 8

RESULT 10

S69164
ferredoxin a1 - Japanese radish (fragments)
C:Species: Kaikware daikon (Japanese radish)
C>Date: 10-Mar-1998 #sequence_revision 10-Mar-1998 #text_change 17-Apr-1998
C:Accession: S69164
R:Obata, S.; Nishimura, M.; Nagai, K.; Sakihama, N.; Shin, M.
Arch. Biochem. Biophys. 316, 797-802, 1995
A:Title: Four ferredoxins from Japanese radish leaves.
A:Reference number: S69164; PMID:95168867; PMID:7864635
A:Accession: S69164
A:Molecule type: protein
A:Residues: 1-17 <OBA>
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein

Query Match 25.3%; Score 21; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TYKV 11
:|:|:
Db 2 TYKV 5

RESULT 11

S63493
disimilatory sulfite reductase gamma chain, membrane-bound and soluble - Desulfovibrio
C:Species: Desulfovibrio desulfuricans
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S63493; S63494
R:Steuber, J.; Arendsen, A.F.; Hagen, W.R.; Kroneck, P.M.H.
Eur. J. Biochem. 233, 873-879, 1995
A:Title: Molecular properties of the dissimilatory sulfite reductase from Desulfovibrio
A:Reference number: S63489; PMID:96085152; PMID:8521853
A:Accession: S63493

A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-8 <STE>
A:Accession: S63494
A>Status: preliminary

A:Molecule type: protein
A:Residues: 1-8 <ST2>

Query Match 24.1%; Score 20; DB 2; Length 8;
Best Local Similarity 60.0%; Pred. No. 2.8e+05;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 QMTYK 10
:|:|:
Db 2 EITYK 6

RESULT 12

G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C:Accession: G41946
R:Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A:Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma ger
A:Reference number: A41946; PMID:92049316; PMID:1658619
A:Accession: G41946
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-9 <WHE>
C:Keywords: T-cell receptor

Query Match 24.1%; Score 20; DB 2; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 TYKVMSG 15
:|:|:|:
Db 1 SYGSYSSG 8

RESULT 13

G85802
hypothetical protein 22947 [imported] - Escherichia coli (strain O157:H7, substrain EDL93;
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85802
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
illet, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: G85802
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-9 <STO>
A:Cross-references: GB:AR005174; NID:gl2515957; PIDN:AAG56883.1; GSPDB:GN00145; UWGP:Z294;
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2947

Query Match 24.1%; Score 20; DB 2; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 MTYKVYMS 14
:|:|:|:
Db 1 MTYTFMLS 8

RESULT 14

S66458
ferredoxin - Rhizobium meliloti (fragment)
C:Species: Rhizobium meliloti
C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S66458
R:Riedel, K.U.; Jouanneau, Y.; Masepohl, B.; Puehler, A.; Klipp, W.

Eur. J. Biochem. 231, 742-746, 1995
 A;Title: A Rhizobium meliloti ferredoxin (Fd_xN) purified from Escherichia coli donates e
 A;Reference number: S66458; MUID:95377307; PMID:7649175
 A;Accession: S66458
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 1-10 <RIE>
 C;Genetics:
 A;Gene: fdxN

Query Match 24.1%; Score 20; DB 2; Length 10;
 Best Local Similarity 57.1%; Pred. No. 2.6e+03;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 VIAQMT 8
 :|||
 Db 4 IIASQXT 10

RESULT 15
 H41946
 T-cell receptor gamma chain (5t.12) - mouse (fragment)
 C;Species: Mus musculus (house mouse)
 C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
 C;Accession: H41946
 R;Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
 Mol. Cell. Biol. 11, 5902-5909, 1991
 A;Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma ge
 A;Reference number: A41946; MUID:92049316; PMID:1659619
 A;Accession: H41946
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-12 <WHE>
 C;Keywords: T-cell receptor

Query Match 24.1%; Score 20; DB 2; Length 12;
 Best Local Similarity 80.0%; Pred. No. 3.1e+03;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 VYMSG 15
 :|||
 Db 7 VYSSG 11

Search completed: August 12, 2004, 07:05:39
 Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:55:55 ; Search time 32 Seconds
(without alignments)
27.662 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKYVMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 943

Minimum DB seq length: 0
Maximum DB seq length: 17

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	22	26.5	9	1	ULAD_HUMAN
2	20	24.1	15	1	MP2A_ORYSA
3	20	24.1	15	1	PLAS_MICAE
4	19	22.9	14	1	ADPA_TENMO
5	19	22.9	15	1	MALT_BACTQ
6	19	22.9	16	1	PZ21_TRIST
7	18	21.7	13	1	ADFB_TENMO
8	18	21.7	15	1	ACEA_ACICA
9	18	21.7	16	1	ODPB_SOLTU
10	17	20.5	11	1	ES1_RAT
11	17	20.5	15	1	COX1_THUOB
12	16	19.3	7	1	WR22_ACHFU
13	16	19.3	8	1	CPD1_ENTFA
14	16	19.3	11	1	NXSN_PSETE
15	16	19.3	11	1	UXB2_YEAST
16	16	19.3	13	1	CRBL_VESLE
17	16	19.3	13	1	CRBL_VESXA
18	16	19.3	14	1	SMS1_MYOSC
19	16	19.3	14	1	SMS_ALIMI
20	16	19.3	15	1	GTS_ASADI
21	16	19.3	15	1	NXSO_PSETE
22	16	19.3	15	1	PLAC_SHEEP
23	16	19.3	15	1	UC27_MAIZE
24	16	19.3	16	1	HBD_CLOPA
25	16	19.3	17	1	TL09_SPIOL
26	15	18.1	17	1	AKH_TABAT
27	15	18.1	8	1	RTF2_PERAM
28	15	18.1	8	1	RT34_BOVIN
29	15	18.1	9	1	NEUX_HUMAN
30	15	18.1	10	1	HTF2_CARMO
31	15	18.1	10	1	HTF_HELZE
32	15	18.1	10	1	HTF_TABAT
33	15	18.1	12	1	CALM_TETTH

34	15	18.1	12	1	PPK4_PERFU
35	15	18.1	14	1	MAST_VESMA
36	15	18.1	14	1	PHI_FRUSE
37	15	18.1	15	1	GR78_HORSE
38	15	18.1	15	1	LEC3_AXIPO
39	15	18.1	15	1	RI3A_SPIOL
40	15	18.1	15	1	RKGG_CARCR
41	15	18.1	15	1	UC30_MAIZE
42	15	18.1	16	1	AU26_LITFA
43	15	18.1	16	1	BRB_BASAL
44	15	18.1	16	1	MMPX_SOLTU
45	15	18.1	16	1	PH2_FRUSE

ALIGNMENTS

```
RESULT 1
ULAD_HUMAN
ID ULAD_HUMAN STANDARD; PRT; 9 AA.
AC P31929;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Unknown protein from 2D-page of liver tissue (Spot 106) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC TISSUE=Liver;
EX MEDLINE=94147969; PubMed=8313870;
RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
RA Tissot J.-D., Bairoch A., Appel R.D., Hochstrasser D.P.;
RT "Human liver protein map: update 1993.";
RL Electrophoresis 14:1216-1222(1993).
CC -!- MISCELLANEOUS: On the 2D-gel the determined pI of this unknown
CC protein is: 6, its MW is: 15 kDa.
DR SWISS-2DPAGE; P31929; HUMAN.
FT NON TER 9
SQ SEQUENCE 9 AA; 1129 MW; D02DFB41B6D3322 CRC64;
Query Match 26.5%; Score 22; DB 1; Length 9;
Best Local Similarity 33.3%; Pred. No. 1.4e+05;
Matches 3; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 3 IAKQMTYKV 11
Db 1 LVKKQTYHI 9
: | :
: | :
: | :
RESULT 2
MP2A_ORYSA
ID MP2A_ORYSA STANDARD; PRT; 15 AA.
AC P83466;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pollen allergen Ory s 2-A (Fragment).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE.
RC STRAIN=cv. Japonica; TISSUE=Pollen;
RA Kerim T., Imin N., Weinman J.J., Rolfe B.G.;
RL Submitted (SEP-2002) to Swiss-Prot.
CC -!- ALLERGEN: Causes an allergic reaction in human. Causes grass
CC pollen allergy. Binds IGE.
CC -!- SIMILARITY: Belongs to the expansin family.
```


OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Viperidae; Crotalinae; Trimeresurus.
 OC NCBI_TaxID=39682;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RA Li S.-Y., Guo Z.-X., Yang Y.-Y., Wang W.-Y., Xiong Y.-L.;
 RT "Isolation and sequencing of five variants of phospholipase A2
 RL from venom of snake Trimeresurus stejnegeri.";
 RL J. Hubei Univ. 25:63-68(2003).
 CC -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
 CC acyl groups in 3-sn-phosphoglycerides.
 CC -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
 CC acylglycerophosphocholine + a fatty acid anion.
 CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -!- SIMILARITY: Belongs to the phospholipase A2 family. Group II
 CC subfamily.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR PROSITE: PS00119; PA2_ASP; PARTIAL.
 DR PROSITE: PS00118; PA2_HIS; PARTIAL.
 KW Hydrolase; Lipid degradation; Calcium; Multigene family.
 FT NON_TER 16
 SQ SEQUENCE 16 AA; 2012 MW; 4EF2D4959E981117 CRC64;
 Query Match 22.9%; Score 19; DB 1; Length 16;
 Best Local Similarity 66.7%; Pred. No. 3.4e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 5 KQMTYK 10
 Db 10 KKMINK 15
 RESULT 7
 ADFB TENMO
 ID ADFB TENMO STANDARD; PRT; 13 AA.
 AC P83109;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Antidiuretic factor B (ADFB).
 OS Tenebrio molitor (Yellow mealworm).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
 OC Tenebrionidae; Tenebrio.
 OK NCBI_TaxID=7067;
 RN [1]
 RP SEQUENCE, FUNCTION, TISSUE SPECIFICITY, MASS SPECTROMETRY, AND
 RP SYNTHESIS.
 RC TISSUE=Head;
 RX MEDLINE=22465067; PubMed=12576082;
 RA Eigenheer R.A., Wierhart U.M., Nicolson S.W., Schoofs L., Schegg K.M.,
 RA Hull J.J., Schooley D.A.;
 RT "Isolation, identification and localization of a second beetle
 RT antidiuretic peptide.";
 RL Peptides 24:27-34(2003).
 CC -!- FUNCTION: Inhibitor of fluid secretion by Malpighian tubules. Uses
 CC CGMP as second messenger. May function as an antidiuretic
 CC hormone.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Brain. Appears to be mainly expressed in two
 CC pairs of bilaterally symmetrical cells in the protocerebrum.
 CC -!- MASS SPECTROMETRY: MW=1560.33; METHOD=MALDI.
 KW Neuropeptide; Hormone.
 SQ SEQUENCE 13 AA; 1562 MW; 0240A4504B8A632B CRC64;
 Query Match 21.7%; Score 18; DB 1; Length 13;
 Best Local Similarity 37.5%; Pred. No. 4.2e+03;
 Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Qy 8 TYKVMSG 15

Db 5 SYKPHYG 12
 RESULT 8
 ACEA ACICA
 ID ACEA ACICA STANDARD; PRT; 15 AA.
 AC P28467;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)
 DE (Fragment).
 GN ACEA.
 OS Acinetobacter calcoaceticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 OK NCBI_TaxID=471;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92041568; PubMed=1938889;
 RA Hoyt J.C., Johnson K.E., Reeves H.C.;
 RT "Purification and characterization of Acinetobacter calcoaceticus
 RT isocitrate lyase.";
 RL J. Bacteriol. 173:6844-6848(1991).
 CC -!- CATALYTIC ACTIVITY: Isocitrate = succinate + glyoxylate.
 CC -!- PATHWAY: Glyoxylate bypass; first step.
 CC -!- SUBUNIT: Homotetramer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the isocitrate lyase family.
 DR PIR; A41338; A41338.
 DR InterPro: IPR000918; Isocit lyase ph.
 DR PROSITE: PS00161; ISOCITRATE_LYASE; PARTIAL.
 KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase.
 FT NON_TER 15
 SQ SEQUENCE 15 AA; 1710 MW; 83AE726B1F2F96E3 CRC64;
 Query Match 21.7%; Score 18; DB 1; Length 15;
 Best Local Similarity 75.0%; Pred. No. 4.8e+03;
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 7 MTKYK 10
 Db 1 MTQYK 4
 RESULT 9
 ODPB SOLTU
 ID ODPB SOLTU STANDARD; PRT; 16 AA.
 AC P81419;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pyruvate dehydrogenase E1 component beta subunit, mitochondrial
 DE (EC 1.2.4.1) (PDHE1-B) (Fragment).
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OK NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=cv. Romano; TISSUE=Tuber;
 RX MEDLINE=98399821; PubMed=9729464;
 RA Millar A.H., Knorr C., Leaver C.J., Hill S.A.;
 RT "Plant mitochondrial pyruvate dehydrogenase complex: purification and
 RT identification of catalytic components in potato.";
 RL Biochem. J. 334:571-576(1998).
 CC -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
 CC conversion of pyruvate to acetyl-CoA and CO(2). It contains
 CC multiple copies of three enzymatic components: pyruvate
 CC dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
 CC lipoamide dehydrogenase (E3).

```

CC -!- CATALYTIC ACTIVITY: Pyruvate + lipoamide = S-
CC acetyldihydrolipoamide + Co(2).
CC -!- COPACTOR: Thiamine pyrophosphate.
CC -!- SUBUNIT: Tetramer of two alpha and two beta subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
KW Phosphorylation; Mitochondrion.
FT NON_TER 16 16
SQ SEQUENCE 16 AA; 1705 MW; FF6ED80EC804F797 CRC64;

Query Match 21.7%; Score 18; DB 1; Length 16;
Best Local Similarity 50.0%; Pred. No. 5.1e+03;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KQMTYK 10
Db 5 KEMTVR 10

RESULT 10
ES1_RAT STANDARD; PRT; 11 AA.
AC P56571;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ES1 protein, mitochondrial (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1];
RP SEQUENCE.
RC STRAIN=Wiistar; TISSUE=Heart;
RA Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA Jungblut P.R.;
RL Submitted (SEP-1998) to Swiss-Prot.
CC -!- SUBCELLULAR LOCATION: Mitochondrial (Potential).
CC -!- MISCELLANEOUS: By 2D-PAGE, the determined pI of this protein (spot
CC P2) is: 8.9, its MW is: 25 kDa.
CC -!- SIMILARITY: BELONGS TO THE ES1 FAMILY.
KW Mitochondrion.
FT NON_TER 11 11
SQ SEQUENCE 11 AA; 1142 MW; D862272D32C72DC2 CRC64;

Query Match 20.5%; Score 17; DB 1; Length 11;
Best Local Similarity 75.0%; Pred. No. 5.4e+03;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 13 MSGT 16
Db 8 LSGT 11

RESULT 11
COXI_THUOB STANDARD; PRT; 15 AA.
AC P80978;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytochrome c oxidase polypeptide Vic-2 (EC 1.9.3.1) (Fragments).
OS Thunus obesus (Bigeye tuna).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8241;
RN [1];
RP SEQUENCE.
RC TISSUE=Heart;
RX MEDLINE=97454291; PubMed=9310366;

```

```

RA Arnold S., Lee I., Kim M., Song E., Linder D., Lottspeich F.,
RA Kadenbach B.;
RT "The subunit structure of cytochrome-c oxidase from tuna heart and
RT liver.";
RL Eur. J. Biochem. 248:99-103(1997).
CC -!- FUNCTION: This protein is one of the nuclear-coded polypeptide
CC chains of cytochrome c oxidase, the terminal oxidase in
CC mitochondrial electron transport.
CC -!- CATALYTIC ACTIVITY: 4 ferrocycochrome c + O(2) = 4 ferricytochrome
CC c + 2 H(2)O.
CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
DR PIR; S77987; S77987.
KW Oxidoreductase; Inner membrane; Mitochondrion.
FT NON_TER 1 1
FT NON_CONS 8 9
FT NON_TER 15 15
SQ SEQUENCE 15 AA; 1696 MW; 4C4C966C73A40294 CRC64;

Query Match 20.5%; Score 17; DB 1; Length 15;
Best Local Similarity 60.0%; Pred. No. 7.2e+03;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VIAKQ 6
Db 5 VVAKK 9

RESULT 12
WWA2_ACHFU STANDARD; PRT; 7 AA.
AC P35920;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Wamide-2.
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Sigmurethra; Achatinoidea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1];
RP SEQUENCE.
RC TISSUE=Ganglion;
RX MEDLINE=93265912; PubMed=8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from
RT ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
DR PIR; S33246; S33246.
KW Neuropeptide; Amidation.
FT MOD_RES 7 7
SQ SEQUENCE 7 AA; 964 MW; 7362D5B686D32310 CRC64;

Query Match 19.3%; Score 16; DB 1; Length 7;
Best Local Similarity 75.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 KQMT 8
Db 2 KQMS 5

RESULT 13
CPD1_ENTFA STANDARD; PRT; 8 AA.
AC P13269;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Sex pheromone cpd1.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1];

```


RP SEQUENCE.
 RX MEDLINE=85040388; PubMed=6436978;
 RA Suzuki A., Mori M., Sagakami Y., Isogai A., Fujino M., Kitada C.,
 RA Craig R.A., Clewell D.B.;
 RT "Isolation and structure of bacterial sex pheromone, cPD1.";
 RL Science 226:849-850(1984).
 CC -!- FUNCTION: cPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
 CC BACTERIOCCIN PLASMID PPDI.
 KW Pheromone.
 SQ SEQUENCE 8 AA; 913 MW; 8665B729C682C729 CRC64;

Query Match 19.3%; Score 16; DB 1; Length 8;
 Best Local Similarity 40.0%; Pred. No. 1.4e+05;
 Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 11 YKVG 15
 DB 4 YKVG 8

RESULT 14
 NXSN PSETE STANDARD; PRT; 11 AA.
 AC P59072;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Short neurotoxin N1 (Alpha neurotoxin) (Fragment).
 OS Pseudonaja textilis (Eastern brown snake).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 CC Elapidae; Acanthophiinae; Pseudonaja.
 OX NCBI_TaxID=8673;
 RN [1]
 RP SEQUENCE, AND MASS SPECTROMETRY.
 RC TISSUE=Venom;
 RX MEDLINE=99449602; PubMed=10518793;
 RA Gong N.L., Armugam A., Jeyaseelan K.;
 RT "Postsynaptic short-chain neurotoxins from Pseudonaja textilis: cDNA
 RT cloning, expression and protein characterization.";
 RL Eur. J. Biochem. 265:982-989(1999).
 CC -!- FUNCTION: Lethal neurotoxin, binds and inhibits nicotinic
 CC acetylcholine receptors (nAChR).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -!- MASS SPECTROMETRY: MW=6236; METHOD=Electrospray.
 CC -!- MISCELLANEOUS: LD(50) is 0.84 mg/kg by intravenous injection.
 CC -!- SIMILARITY: Belongs to the snake toxin family.
 DR InterPro; IPR003571; Snake_toxin.
 DR PROSITE; PS00272; SNAKE_TOXIN; PARTIAL.
 KW Toxin; Neurotoxin; Postsynaptic neurotoxin;
 KW Acetylcholine receptor inhibitor; Multigene family.
 FT UNSURE 3
 FT NON_TER 11 11
 SQ SEQUENCE 11 AA; 1319 MW; 0D1EF0C81B59732B CRC64;

Query Match 19.3%; Score 16; DB 1; Length 11;
 Best Local Similarity 75.0%; Pred. No. 8.1e+03;
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 YKVG 12
 DB 4 YKVG 7

RESULT 15
 UXB2 YEAST STANDARD; PRT; 11 AA.
 AC P99013;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Unknown protein from 2D-page (Spot 2D-000K2F) (Fragment).

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=X2180-1A;
 RA Sanchez J.-C., Golaz O., Schaller D., Morch F., Frutiger S.,
 RA Hughes G.J., Appel R.D., Deshusses J., Hochstrasser D.F.;
 RL Submitted (AUG-1995) to Swiss-Prot.
 CC -!- MISCELLANEOUS: On the 2D-gel the determined pI of this unknown
 CC protein is: 6.20, its MW is: 9.2 kDa.
 DR SWISS-2DPAGE; P99013; YEAST.
 FT NON_TER 11 11
 SQ SEQUENCE 11 AA; 1328 MW; EC38021C0DCB42DA CRC64;

Query Match 19.3%; Score 16; DB 1; Length 11;
 Best Local Similarity 100.0%; Pred. No. 8.1e+03;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 KVG 12
 DB 9 KVG 11

Search completed: August 12, 2004, 07:04:05
 Job time : 33 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 07:01:00 ; Search time 34 seconds
(without alignments)
157.759 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVMGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 5061

Minimum DB seq length: 0
Maximum DB seq length: 17

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phage:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	23	27.7	13	13 Q8QZ5	Q8qz5 fugu rubrip
2	23	27.7	15	8 Q9XIJ8	Q9xij8 grus leucog
3	23	27.7	15	8 Q9TH03	Q9th03 grus paradi
4	22	26.5	13	5 Q8T6E9	Q8t6e9 drosophila
5	22	26.5	13	5 Q8STI5	Q8sti5 drosophila
6	22	26.5	13	5 Q35793	Q35793 saccharomyc
7	21	25.3	8	7 Q9S213	Q9s213 oryctolagus
8	21	25.3	12	2 Q9R3B3	Q9r3b3 helicobacte
9	21	25.3	13	2 Q8GL29	Q8gl29 borrelia bu
10	21	25.3	17	2 Q9R575	Q9r575 synechocyst
11	20	24.1	9	16 Q8X4G1	Q8x4gl escherichia
12	20	24.1	11	7 Q77900	Q77900 oreochromis
13	20	24.1	11	7 Q77917	Q77917 oreochromis
14	20	24.1	11	7 Q77902	Q77902 oreochromis
15	20	24.1	11	7 Q77921	Q77921 pseudotroph
16	20	24.1	11	7 Q77901	Q77901 oreochromis

Q77916 oreochromis
Q77905 oreochromis
Q77899 oreochromis
Q77904 oreochromis
Q77903 oreochromis
Q9bft9 tupaia mino
Q9ge05 etaballia g
Q9afk8 brugia paba
Q8m099 tockus nasu
Q7x8p6 zea mays (m
O52059 salmonella
O05991 staphylococ
Q7xb01 zea mays (m
Q7xb00 zea mays (m
Q9Sm49 bos taurus
Q7xaz9 zea mays (m
P83158 anabaena sp
Q9byy9 homo sapien
Q9bft6 ateles fusc
Q9bfv6 chaetophrac
Q9bfr6 canis famill
Q9bfs3 okapia john
Q9bfr1 condylura c
Q9bft7 tarsius ban
Q9bfr4 manis penta
Q9bfs9 megaptera n
Q9bfs3 erinaceus c
Q9bfs5 tamandua te
Q9bfs5 tragelaphus

ALIGNMENTS

RESULT 1

Q8QZ5 PRELIMINARY; PRT; 13 AA.
AC Q8QZ5;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Guanine nucleotide binding protein (Fragment).
GN GNAO.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97129408; PubMed=8973916;
RA Sarwal M.M., Sontag J.M., Hoang L., Brenner S., Wilkie T.M.;
RT "G protein alpha subunit multigene family in the Japanese puffer fish
Fugu rubripes: PCR from a compact vertebrate genome.";
RL Genome Res. 6:1207-1215(1996).
DR EMBL; I79891; AAL7640.1;
DR GO; GO:0004874; P:signal transducer activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR001019; Gprotein_alpha.
DR Pfam; PF00503; G-alpha; 1.
FT NON_TER 1
FT NON_TER 13
SQ SEQUENCE 13 AA; 1336 MW; 465859640B44B5B3 CRC64;

Query Match 27.7%; Score 23; DB 13; Length 13;

Best Local Similarity 71.4%; Pred. No. 3.6e+03;

Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SVIAKQM 7

Db 7 STIVKQM 13

```
RESULT 2
Q9XLJ8
ID Q9XLJ8 PRELIMINARY; PRT; 15 AA.
AC Q9XLJ8;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE NADH dehydrogenase subunit 6 (Fragment).
GN ND6.
OS Grus leucogeranus (Siberian crane).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Gruiformes; Gruidae; Grus.
OX NCBI_TaxID=40819;
RN [1]
RP SEQUENCE FROM N.A.
RA Glenn T.C., Stephan W., Braun M.J.;
RT "Effects of a Population Bottleneck on Whooping Crane Mitochondrial
RT DNA Variation.";
RL Conserv. Biol. 0:0-0(1999).
DR EMBL: AF112371; AAD23992.1; -.
DR GO: GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON_TER 15
SQ SEQUENCE 15 AA; 1807 MW; A7FFB3A65E8A734F CRC64;

Query Match 27.7%; Score 23; DB 8; Length 15;
Best Local Similarity 44.4%; Pred. No. 4.2e+03;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 MTKYKVMMSG 15
||| : :
Db 1 MTKYFFFMG 9

RESULT 3
Q9TH03
ID Q9TH03 PRELIMINARY; PRT; 15 AA.
AC Q9TH03;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE NADH dehydrogenase subunit 6 (Fragment).
GN ND6.
OS Grus paradisea (Blue crane).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Gruiformes; Gruidae; Grus.
OX NCBI_TaxID=40825;
RN [1]
RP SEQUENCE FROM N.A.
RA Glenn T.C., Stephan W., Braun M.J.;
RT "Effects of a Population Bottleneck on Whooping Crane Mitochondrial
RT DNA Variation.";
RL Conserv. Biol. 0:0-0(1999).
DR EMBL: AF112372; AAD23993.1; -.
DR GO: GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON_TER 15
SQ SEQUENCE 15 AA; 1807 MW; A7FFB3A65E8A734F CRC64;

Query Match 27.7%; Score 23; DB 8; Length 15;
Best Local Similarity 44.4%; Pred. No. 4.2e+03;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 MTKYKVMMSG 15
||| : :
Db 1 MTKYFFFMG 9

RESULT 4
Q9T6E9
ID Q9T6E9 PRELIMINARY; PRT; 13 AA.
AC Q9T6E9;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Myocyte enhancing factor 2 (Fragment).
GN MEF2.
OS Drosophila miranda (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7229;
RN [1]
RP SEQUENCE FROM N.A.
RA Schaeffer S.W., Goetting-Minesky M.P., Kovacevic M., Peoples J.,
RA Graybill J.L., Miller J.M., Kim K., Nelson J.G., Anderson W.W.;
RT "Evolutionary genomics of inversions in Drosophila pseudoobscura:
RT Modes of selection.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF476817; AAL91815.1; -.
DR FlyBase; FBgn0062329; Dmir\Mef2.
FT NON_TER 1
FT NON_TER 13
SQ SEQUENCE 13 AA; 1313 MW; 15BF380B6BE05050 CRC64;

Query Match 26.5%; Score 22; DB 5; Length 13;
Best Local Similarity 36.4%; Pred. No. 5.4e+03;
Matches 4; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 7 MTKYKVMMSGTV 17
| : | | | :
Db 3 MSLIIPSGSM 13

RESULT 5
Q8STI5
ID Q8STI5 PRELIMINARY; PRT; 13 AA.
AC Q8STI5;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Myocyte enhancing factor 2 (Fragment).
GN MEF2.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7237;
RN [1]
RP SEQUENCE FROM N.A.
RA Schaeffer S.W., Goetting-Minesky M.P., Kovacevic M., Peoples J.,
RA Graybill J.L., Miller J.M., Kim K., Nelson J.G., Anderson W.W.;
RT "Evolutionary genomics of inversions in Drosophila pseudoobscura:
RT Modes of selection.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF476728; AAL91726.1; -.
DR EMBL: AF476729; AAL91727.1; -.
DR EMBL: AF476730; AAL91728.1; -.
DR EMBL: AF476731; AAL91729.1; -.
DR EMBL: AF476732; AAL91730.1; -.
DR EMBL: AF476733; AAL91731.1; -.
DR EMBL: AF476734; AAL91732.1; -.
DR EMBL: AF476735; AAL91733.1; -.
DR EMBL: AF476736; AAL91734.1; -.
DR EMBL: AF476737; AAL91735.1; -.
DR EMBL: AF476738; AAL91736.1; -.
DR EMBL: AF476739; AAL91737.1; -.
DR EMBL: AF476740; AAL91738.1; -.
DR EMBL: AF476741; AAL91739.1; -.
DR EMBL: AF476742; AAL91740.1; -.
DR EMBL: AF476743; AAL91741.1; -.
DR EMBL: AF476744; AAL91742.1; -.
DR EMBL: AF476745; AAL91743.1; -.

```

```
DR EMBL; AF476746; AAL91744.1; --
DR EMBL; AF476747; AAL91745.1; --
DR EMBL; AF476748; AAL91746.1; --
DR EMBL; AF476749; AAL91747.1; --
DR EMBL; AF476750; AAL91748.1; --
DR EMBL; AF476751; AAL91749.1; --
DR EMBL; AF476752; AAL91750.1; --
DR EMBL; AF476753; AAL91751.1; --
DR EMBL; AF476754; AAL91752.1; --
DR EMBL; AF476755; AAL91753.1; --
DR EMBL; AF476756; AAL91754.1; --
DR EMBL; AF476757; AAL91755.1; --
DR EMBL; AF476758; AAL91756.1; --
DR EMBL; AF476759; AAL91757.1; --
DR EMBL; AF476760; AAL91758.1; --
DR EMBL; AF476761; AAL91759.1; --
DR EMBL; AF476762; AAL91760.1; --
DR EMBL; AF476763; AAL91761.1; --
DR EMBL; AF476764; AAL91762.1; --
DR EMBL; AF476765; AAL91763.1; --
DR EMBL; AF476766; AAL91764.1; --
DR EMBL; AF476767; AAL91765.1; --
DR EMBL; AF476768; AAL91766.1; --
DR EMBL; AF476769; AAL91767.1; --
DR EMBL; AF476770; AAL91768.1; --
DR EMBL; AF476771; AAL91769.1; --
DR EMBL; AF476772; AAL91770.1; --
DR EMBL; AF476773; AAL91771.1; --
DR EMBL; AF476774; AAL91772.1; --
DR EMBL; AF476775; AAL91773.1; --
DR EMBL; AF476776; AAL91774.1; --
DR EMBL; AF476777; AAL91775.1; --
DR EMBL; AF476778; AAL91776.1; --
DR EMBL; AF476779; AAL91777.1; --
DR EMBL; AF476780; AAL91778.1; --
DR EMBL; AF476781; AAL91779.1; --
DR EMBL; AF476782; AAL91780.1; --
DR EMBL; AF476783; AAL91781.1; --
DR EMBL; AF476784; AAL91782.1; --
DR EMBL; AF476785; AAL91783.1; --
DR EMBL; AF476786; AAL91784.1; --
DR EMBL; AF476787; AAL91785.1; --
DR EMBL; AF476788; AAL91786.1; --
DR EMBL; AF476789; AAL91787.1; --
DR EMBL; AF476790; AAL91788.1; --
DR EMBL; AF476791; AAL91789.1; --
DR EMBL; AF476792; AAL91790.1; --
DR EMBL; AF476793; AAL91791.1; --
DR EMBL; AF476794; AAL91792.1; --
DR EMBL; AF476795; AAL91793.1; --
DR EMBL; AF476796; AAL91794.1; --
DR EMBL; AF476797; AAL91795.1; --
DR EMBL; AF476798; AAL91796.1; --
DR EMBL; AF476799; AAL91797.1; --
DR EMBL; AF476800; AAL91798.1; --
DR EMBL; AF476801; AAL91799.1; --
DR EMBL; AF476802; AAL91800.1; --
DR EMBL; AF476803; AAL91801.1; --
DR EMBL; AF476804; AAL91802.1; --
DR EMBL; AF476805; AAL91803.1; --
DR EMBL; AF476806; AAL91804.1; --
DR EMBL; AF476807; AAL91805.1; --
DR EMBL; AF476808; AAL91806.1; --
DR EMBL; AF476809; AAL91807.1; --
DR EMBL; AF476810; AAL91808.1; --
DR EMBL; AF476811; AAL91809.1; --
DR EMBL; AF476812; AAL91810.1; --
DR EMBL; AF476813; AAL91811.1; --
DR EMBL; AF476814; AAL91812.1; --
DR EMBL; AF476815; AAL91813.1; --
DR EMBL; AF476816; AAL91814.1; --
DR FlyBase; FBgn0062298; Dpse\Mef2.
FT NON_TER 1 1
```

```
FT NON_TER 13
SQ SEQUENCE 13 AA; 1313 MW; 15BF380B6BE05050 CRC64;

Query Match 26.5%; Score 22; DB 5; Length 13;
Best Local Similarity 36.4%; Pred. No. 5.4e+03;
Matches 4; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 MTYKVYMSGTV 17
  | : | : | :
Db 3 MSLIYPSGSM 13

RESULT 6
Q35793 PRELIMINARY; PRT; 13 AA.
ID Q35793;
AC Q35793;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Inside intron 4 (Fragment).
OS Saccharomyces cerevisiae (Baker's yeast).
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=D273-10B;
RX MEDLINE=81069885; PubMed=6254986;
RA Bonitz S.G., Coruzzi G., Thalenfeld B., Tzagoloff A., Macino G.;
RT "Assembly of the mitochondrial membrane system: Structure and
RT nucleotide sequence of the gene coding for subunit 1 of yeast
RT cytochrome oxidase.";
RL J. Biol. Chem. 255:11927-11941(1980).
DR EMBL; V00694; CAA24065.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 13 13
SQ SEQUENCE 13 AA; 1749 MW; 114378726A89945B CRC64;

Query Match 26.5%; Score 22; DB 8; Length 13;
Best Local Similarity 40.0%; Pred. No. 5.4e+03;
Matches 4; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 AKQTYKVYM 13
  : | | : | |
Db 1 SKLYMNNYM 10

RESULT 7
Q95213 PRELIMINARY; PRT; 8 AA.
ID Q95213;
AC Q95213;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Germline DH (Df) gene (Fragment).
GN DF.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F-1/rgm;
RA Mage R.G., Chen H.-T., Alexander C.B., Chen P.F.;
RT "Rabbit DQ52 and DH Gene Rearrangements in Early B-cell Development.";
RL Mol. Immunol. 0:0-0(1996).
DR EMBL; U62585; AAB18735.1; -.
FT NON_TER 1 1
FT NON_TER 8 8
SQ SEQUENCE 8 AA; 845 MW; 5CA861B5AB58677B CRC64;
```

Query Match 25.3%; Score 21; DB 7; Length 8;
 Best Local Similarity 50.0%; Pred. No. 1e+06;
 Matches 4; Conservative 1; Mismatches 0; Gaps 0;

Qy 9 YKVTMSGT 16
 Db 1 YPGYSTGT 8

RESULT 8
 ID Q9R3B3 PRELIMINARY; PRT; 12 AA.
 AC Q9R3B3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Urace large subunit (Fragment).
 OS Helicobacter felis, and
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=214, 210;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=93084378; PubMed=1452359;
 RA Turbett G.R., Hoj P.B., Horne R., Mee B.J.;
 RT "Purification and characterization of the urease enzymes of
 RT Helicobacter species from humans and animals."
 RL Infect. Immun. 60:5259-5266 (1992).
 DR PIR; C49215; C49215.
 SQ SEQUENCE 12 AA; 1500 MW; 93F05D2362132415 CRC64;

Query Match 25.3%; Score 21; DB 2; Length 12;
 Best Local Similarity 40.0%; Pred. No. 7.5e+03;
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KQTKYKVMYS 14
 Db 2 KKISKEYVS 11

RESULT 9
 ID Q8GL29 PRELIMINARY; PRT; 13 AA.
 AC Q8GL29;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Bdr protein (Fragment).
 GN BDR.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OG Plasmid group cp32-1.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sh-2-82;
 RA Stevenson B., Miller J.C.;
 RT "Comparative analyses of Borrelia burgdorferi erp genes and their cp32
 RT phages: conservation amidst diversity."
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY142089; AANL7872.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Plasmid.
 FT NON_TER 13 13
 SQ SEQUENCE 13 AA; 1548 MW; 08E030E20078A32B CRC64;

Query Match 25.3%; Score 21; DB 2; Length 13;
 Best Local Similarity 50.0%; Pred. No. 8.2e+03;
 Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 7 MTYKVY 12

Db 4 LAYKTY 9

RESULT 10
 ID Q9R575 PRELIMINARY; PRT; 17 AA.
 AC Q9R575;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE NAD(P)H-plastoquinone-oxidoreductase 18 kDa polypeptide (Fragment).
 OS Synechocystis.
 OC Bacteria; Cyanobacteria; Chroococcales.
 OX NCBI_TaxID=1142;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=93314795; PubMed=8325373;
 RA Berger S., Ellersiek U., Kinzelt D., Steinmuller K.;
 RT "Immunopurification of a subcomplex of the NAD(P)H-plastoquinone-
 RT oxidoreductase from the cyanobacterium Synechocystis sp. PCC6803."
 RL FEBS Lett. 326:246-250 (1993).
 SQ SEQUENCE 17 AA; 1890 MW; F776D9E9FE2A58FF7 CRC64;

Query Match 25.3%; Score 21; DB 2; Length 17;
 Best Local Similarity 40.0%; Pred. No. 1.1e+04;
 Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 8 TYKVYMSGTV 17
 Db 3 TVKVVLNETI 12

RESULT 11
 ID Q8X4G1 PRELIMINARY; PRT; 9 AA.
 AC Q8X4G1;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Hypothetical protein Z2947.
 GN Z2947.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 DR EMBL; AE005411; AAG56883.1; -.
 DR PIR; G85802; G85802.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 9 AA; 1107 MW; 8F6CB72699D1BB41 CRC64;

Query Match 24.1%; Score 20; DB 16; Length 9;
 Best Local Similarity 50.0%; Pred. No. 1e+06;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 7 MTYKVYMS 14
 Db 1 MTYTFMLS 8

RESULT 12

```
O77900
ID O77900 PRELIMINARY; PRT; 11 AA.
AC O77900;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MHC class II B locus 14 (Fragment).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
OC Cichlidae; Oreochromis.
OX NCBI_TaxID=8128;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., Zaleska-Rutczynska Z., McAndrew B., Vincek V.,
RA Figueroa F., Sultmann H., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid mhc
RT class II B loci.";
RL Genetics 149:1527-1537(1998).
DR EMBL; AF050010; AAC41349.1; -.
FT NON_TER 1
FT NON_TER 11
SQ SEQUENCE 11 AA; 1349 MW; 81C12D8EB7341B41 CRC64;

Query Match 24.1%; Score 20; DB 7; Length 11;
Best Local Similarity 60.0%; Pred. No. 1e+04;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 MTKV 11
Db 3 MTKV 7

RESULT 13
O77917
ID O77917 PRELIMINARY; PRT; 11 AA.
AC O77917;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MHC class II B locus 14 (Fragment).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
OC Cichlidae; Oreochromis.
OX NCBI_TaxID=8128;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., Zaleska-Rutczynska Z., McAndrew B., Vincek V.,
RA Figueroa F., Sultmann H., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid mhc
RT class II B loci.";
RL Genetics 149:1527-1537(1998).
DR EMBL; AF050030; AAC41369.1; -.
FT NON_TER 1
FT NON_TER 11
SQ SEQUENCE 11 AA; 1349 MW; 81C12D8EB7341B41 CRC64;

Query Match 24.1%; Score 20; DB 7; Length 11;
Best Local Similarity 60.0%; Pred. No. 1e+04;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 MTKV 11
Db 3 MTKV 7

RESULT 14
O77902
ID O77902 PRELIMINARY; PRT; 11 AA.
AC O77902;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MHC class II B locus 14 (Fragment).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
OC Cichlidae; Oreochromis.
OX NCBI_TaxID=8128;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., Zaleska-Rutczynska Z., McAndrew B., Vincek V.,
RA Figueroa F., Sultmann H., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid mhc
RT class II B loci.";
RL Genetics 149:1527-1537(1998).
DR EMBL; AF050012; AAC41351.1; -.
FT NON_TER 1
FT NON_TER 11
SQ SEQUENCE 11 AA; 1349 MW; 81C12D8EB7341B41 CRC64;

Query Match 24.1%; Score 20; DB 7; Length 11;
Best Local Similarity 60.0%; Pred. No. 1e+04;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 MTKV 11
Db 3 MTKV 7

RESULT 15
O77921
ID O77921 PRELIMINARY; PRT; 11 AA.
AC O77921;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MHC class II B locus 14 (Fragment).
OS Pseudotropheus sp. 'Pseudotropheus tropheus complex'.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
OC Cichlidae; Pseudotropheus.
OX NCBI_TaxID=51796;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., Zaleska-Rutczynska Z., McAndrew B., Vincek V.,
RA Figueroa F., Sultmann H., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid mhc
RT class II B loci.";
RL Genetics 149:1527-1537(1998).
DR EMBL; AF050034; AAC41373.1; -.
FT NON_TER 1
FT NON_TER 11
SQ SEQUENCE 11 AA; 1349 MW; 81C12D8EB7341B41 CRC64;

Query Match 24.1%; Score 20; DB 7; Length 11;
Best Local Similarity 60.0%; Pred. No. 1e+04;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 MTKV 11
Db 3 MTKV 7

Search completed: August 12, 2004, 07:04:51
Job time : 36 secs
```

This page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 07:04:56 ; Search time 41 Seconds
(without alignments)
130.165 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVMGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 228781

Minimum DB seq length: 0

Maximum DB seq length: 17

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	36.1	11	US-10-360-101-187	Sequence 187, App
2	30	33.7	9	US-10-253-286-492	Sequence 492, App
3	28	33.7	9	US-10-253-286-492	Sequence 493, App
4	28	33.7	9	US-10-245-871-492	Sequence 492, App
5	28	33.7	9	US-10-245-871-492	Sequence 493, App
6	28	33.7	14	US-10-253-286-508	Sequence 508, App
7	28	33.7	14	US-10-253-286-509	Sequence 508, App
8	28	33.7	14	US-10-245-871-508	Sequence 509, App
9	28	33.7	14	US-10-245-871-509	Sequence 508, App
10	28	33.7	17	US-10-253-286-501	Sequence 501, App
11	28	33.7	17	US-10-253-286-510	Sequence 510, App
12	28	33.7	17	US-10-245-871-501	Sequence 501, App
13	28	33.7	17	US-10-245-871-510	Sequence 510, App
14	27	32.5	10	US-10-200-708-291	Sequence 291, App
15	27	32.5	13	US-10-246-581-24	Sequence 24, App

16	26	31.3	10	8	US-08-325-278-12	Sequence 12, Appl
17	26	31.3	10	12	US-10-416-822-4	Sequence 4, Appl
18	26	31.3	12	14	US-10-246-581-22	Sequence 22, Appl
19	26	31.3	13	12	US-09-988-493-253	Sequence 253, Appl
20	26	31.3	14	16	US-10-762-629-48	Sequence 48, Appl
21	26	31.3	16	14	US-10-225-567A-2292	Sequence 2292, Appl
22	26	31.3	17	11	US-09-754-831A-34	Sequence 34, Appl
23	25	30.1	8	15	US-10-388-337-11	Sequence 11, Appl
24	25	30.1	9	9	US-09-812-528-5	Sequence 5, Appl
25	25	30.1	9	9	US-09-847-185-38	Sequence 38, Appl
26	25	30.1	9	9	US-09-923-831-21	Sequence 21, Appl
27	25	30.1	9	9	US-09-872-832-13	Sequence 13, Appl
28	25	30.1	9	9	US-09-888-721-27	Sequence 27, Appl
29	25	30.1	9	9	US-09-888-721-28	Sequence 28, Appl
30	25	30.1	9	9	US-09-766-889A-33	Sequence 33, Appl
31	25	30.1	9	9	US-09-909-460-60	Sequence 60, Appl
32	25	30.1	9	10	US-09-898-860-32	Sequence 32, Appl
33	25	30.1	9	11	US-09-077-439A-7	Sequence 7, Appl
34	25	30.1	9	12	US-10-218-095-28	Sequence 28, Appl
35	25	30.1	9	12	US-10-253-286-519	Sequence 519, Appl
36	25	30.1	9	12	US-09-775-805-12	Sequence 12, Appl
37	25	30.1	9	12	US-09-077-214-11	Sequence 11, Appl
38	25	30.1	9	12	US-10-289-566-1	Sequence 1, Appl
39	25	30.1	9	12	US-10-367-580-145	Sequence 145, App
40	25	30.1	9	12	US-10-367-593-145	Sequence 145, App
41	25	30.1	9	12	US-10-367-594-145	Sequence 145, App
42	25	30.1	9	12	US-10-367-654-145	Sequence 145, App
43	25	30.1	9	12	US-10-367-658-145	Sequence 145, App
44	25	30.1	9	12	US-10-367-668-145	Sequence 145, App
45	25	30.1	9	12	US-09-872-836-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-10-360-101-187
; Sequence 187, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360.101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of interleukin-2 fragment (60-70) (T-cell growth factor)
US-10-360-101-187

Query Match 36.1%; Score 30; DB 15; Length 11;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 7 MTKYKVMYS 14
: : : : :
Db 1 LTFKFKVMS 8

RESULT 2
US-10-253-286-492
; Sequence 492, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT

```
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 492
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-253-286-492

Query Match      33.7%  Score 28; DB 12; Length 9;
Best Local Similarity 57.1%  Pred. No. 1.2e+06;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
Db      3 IYMGTM 9

RESULT 3
US-10-253-286-493
; Sequence 493, Application US/10253286
; Publication No. US2004005881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 493
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-253-286-493

Query Match      33.7%  Score 28; DB 12; Length 9;
Best Local Similarity 57.1%  Pred. No. 1.2e+06;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
Db      3 IYMGTM 9

RESULT 4
US-10-245-871-492
; Sequence 492, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 492
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-245-871-492

Query Match      33.7%  Score 28; DB 12; Length 9;
Best Local Similarity 57.1%  Pred. No. 1.2e+06;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
Db      3 IYMGTM 9

RESULT 5
US-10-245-871-493
; Sequence 493, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 493
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-245-871-493

Query Match      33.7%  Score 28; DB 15; Length 9;
Best Local Similarity 57.1%  Pred. No. 1.2e+06;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
Db      3 IYMGTM 9

RESULT 6
US-10-253-286-508
; Sequence 508, Application US/10253286
; Publication No. US2004005881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 508
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: I1-key/tyrosinase overlapping hybrid peptide
```

```
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 492
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-245-871-492

Query Match      33.7%  Score 28; DB 15; Length 9;
Best Local Similarity 57.1%  Pred. No. 1.2e+06;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
Db      3 IYMGTM 9

RESULT 5
US-10-245-871-493
; Sequence 493, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 493
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-245-871-493

Query Match      33.7%  Score 28; DB 15; Length 9;
Best Local Similarity 57.1%  Pred. No. 1.2e+06;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
Db      3 IYMGTM 9

RESULT 6
US-10-253-286-508
; Sequence 508, Application US/10253286
; Publication No. US2004005881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 508
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: I1-key/tyrosinase overlapping hybrid peptide
```

```
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)
; OTHER INFORMATION: a-aminovaleic acid
; FEATURE:
; OTHER INFORMATION: C-term amidated
US-10-253-286-508

Query Match      33.7%; Score 28; DB 12; Length 14;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      11 VYMSGTV 17
       :||:|:|:
Db       8 IYMGTM 14

RESULT 7
US-10-253-286-509
; Sequence 509, Application US/10253286
; Publication No. US2004005881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPIOTOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 509
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: II-key/tyrosinase overlapping hybrid peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)
; OTHER INFORMATION: a-aminovaleic acid
; FEATURE:
; OTHER INFORMATION: C-term amidated
US-10-253-286-509

Query Match      33.7%; Score 28; DB 12; Length 14;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      11 VYMSGTV 17
       :||:|:|:
Db       6 IYMGTM 12

RESULT 8
US-10-245-871-508
; Sequence 508, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPIOTOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
```

```
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 508
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: II-key/tyrosinase overlapping hybrid peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)
; OTHER INFORMATION: a-aminovaleic acid
; FEATURE:
; OTHER INFORMATION: C-term amidated
US-10-245-871-508

Query Match      33.7%; Score 28; DB 15; Length 14;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      11 VYMSGTV 17
       :||:|:|:
Db       8 IYMGTM 14

RESULT 9
US-10-245-871-509
; Sequence 509, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPIOTOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 509
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: II-key/tyrosinase overlapping hybrid peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)
; OTHER INFORMATION: a-aminovaleic acid
; FEATURE:
; OTHER INFORMATION: C-term amidated
US-10-245-871-509

Query Match      33.7%; Score 28; DB 15; Length 14;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      11 VYMSGTV 17
       :||:|:|:
Db       6 IYMGTM 12

RESULT 10
US-10-253-286-501
; Sequence 501, Application US/10253286
; Publication No. US2004005881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
```

```

; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPI TOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 501
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-253-286-501

Query Match      33.7%; Score 28; DB 12; Length 17;
Best Local Similarity 57.1%; Pred. No. 3.4e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
       :|||:
Db      4 IYNGTM 10

RESULT 11
US-10-253-286-510
; Sequence 510, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPI TOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: II-key/tyrosinase overlapping hybrid peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)
; OTHER INFORMATION: a-aminovaleic acid
; FEATURE:
; OTHER INFORMATION: C-term amidated
; US-10-253-286-510

Query Match      33.7%; Score 28; DB 12; Length 17;
Best Local Similarity 57.1%; Pred. No. 3.4e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
       :|||:
Db      9 IYNGTM 15

RESULT 12
US-10-245-871-501
; Sequence 501, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT

```

```

; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPI TOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 501
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-245-871-501

Query Match      33.7%; Score 28; DB 15; Length 17;
Best Local Similarity 57.1%; Pred. No. 3.4e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
       :|||:
Db      4 IYNGTM 10

RESULT 13
US-10-245-871-510
; Sequence 510, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPI TOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: II-key/tyrosinase overlapping hybrid peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)
; OTHER INFORMATION: a-aminovaleic acid
; FEATURE:
; OTHER INFORMATION: C-term amidated
; US-10-245-871-510

Query Match      33.7%; Score 28; DB 15; Length 17;
Best Local Similarity 57.1%; Pred. No. 3.4e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      11 VYMSGTV 17
       :|||:
Db      9 IYNGTM 15

RESULT 14
US-10-200-708-291
; Sequence 291, Application US/10200708
; Publication No. US20030180314A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S.

```

```

; TITLE OF INVENTION: HIV VACCINE CANDIDATE PEPTIDES
; FILE REFERENCE: 17999-001
; CURRENT APPLICATION NUMBER: US/10/200,708
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/351,036
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/092,346
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/115,145
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 60/130,677
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 291
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-10-200-708-291

```

```

Query Match          32.5%; Score 27; DB 14; Length 10;
Best Local Similarity 57.1%; Pred. No. 2.9e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      6 QWTYKXY 12
Db      4 QWTYQIY 10

```

```

RESULT 15
US-10-246-581-24
; Sequence 24, Application US/10246581
; Publication No. US20030097680A1
; GENERAL INFORMATION:
; APPLICANT: Bowen, Benjamin A.
; APPLICANT: Chamberlin, Mark A.
; APPLICANT: Drummond, Bruce J.
; APPLICANT: McElver, John A.
; APPLICANT: Rothstein, Rodney J.
; TITLE OF INVENTION: RAD51 Polypeptides
; FILE REFERENCE: 0556D
; CURRENT APPLICATION NUMBER: US/10/246,581
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 09/246,963
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: US 60/074743
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein sequence for GFPm to ZmRAD51B fusion,
; OTHER INFORMATION: including isoleucine and histidine linker
US-10-246-581-24

```

```

Query Match          32.5%; Score 27; DB 14; Length 13;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      9 YKVVMSGT 16
Db      4 YKIHSSS 11

```

```

Search completed: August 12, 2004, 07:10:54
Job time : 41 secs

```

This page blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 07:02:25 ; Search time 40 Seconds
(without alignments)
21.941 Million cell updates/sec

Title: US-09-890-463-2

Perfect score: 83

Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 157007

Minimum DB seq length: 0

Maximum DB seq length: 17

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgm2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgm2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgm2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgm2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgm2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgm2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	29	34.9	14	1	US-08-443-104-2
2	29	34.9	14	1	US-08-238-130-3
3	29	34.9	14	1	US-08-442-859-2
4	29	34.9	14	2	US-08-398-489-2
5	29	34.9	14	2	US-08-894-772-3
6	29	34.9	14	2	US-09-207-844-3
7	29	34.9	14	5	PCT-US95-05534-2
8	28	33.7	12	4	US-08-528-200-40
9	28	33.7	13	2	US-08-413-708B-6
10	28	33.7	13	4	US-08-528-200-33
11	28	33.7	13	4	US-09-528-200-39
12	28	33.7	14	4	US-09-528-200-26
13	28	33.7	14	4	US-09-528-200-32
14	28	33.7	15	1	US-08-486-721A-16
15	28	33.7	15	4	US-09-528-200-19
16	28	33.7	15	4	US-09-528-200-25
17	28	33.7	16	4	US-09-528-200-12
18	28	33.7	16	4	US-09-528-200-18
19	28	33.7	17	1	US-07-924-054-7
20	28	33.7	17	4	US-09-528-200-11
21	27	32.5	10	4	US-09-528-200-42
22	27	32.5	11	4	US-09-528-200-35
23	27	32.5	11	4	US-09-528-200-41
24	27	32.5	11	4	US-09-528-200-47
25	27	32.5	11	4	US-09-528-200-190
26	27	32.5	11	4	US-09-528-200-192
27	27	32.5	12	3	US-08-912-560-7

Sequence 28, Appl
Sequence 34, Appl
Sequence 46, Appl
Sequence 191, Appl
Sequence 21, Appl
Sequence 24, Appl
Sequence 21, Appl
Sequence 27, Appl
Sequence 6, Appl
Sequence 23, Appl
Sequence 25, Appl
Sequence 14, Appl
Sequence 20, Appl
Patent No. 5194585
Sequence 13, Appl
Sequence 12, Appl
Sequence 24, Appl
Sequence 12, Appl

28 27 32.5 12 4 US-09-528-200-28
29 27 32.5 12 4 US-09-528-200-34
30 27 32.5 12 4 US-09-528-200-46
31 27 32.5 12 4 US-09-528-200-191
32 27 32.5 13 3 US-08-952-568-21
33 27 32.5 13 4 US-09-246-963A-24
34 27 32.5 13 4 US-09-528-200-21
35 27 32.5 13 4 US-09-528-200-27
36 27 32.5 14 3 US-08-930-845-6
37 27 32.5 14 3 US-08-952-568-23
38 27 32.5 14 3 US-09-192-048-25
39 27 32.5 14 4 US-09-528-200-14
40 27 32.5 14 4 US-09-528-200-20
41 27 32.5 14 6 5194585-13
42 27 32.5 15 4 US-09-528-200-13
43 27 32.5 17 3 US-08-896-933-12
44 27 32.5 17 3 US-09-288-391-24
45 27 32.5 17 4 US-09-314-235-12

ALIGNMENTS

RESULT 1
US-08-443-104-2
; Sequence 2, Application US/08443104
; Patent No. 5691162
; GENERAL INFORMATION:
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; APPLICANT: Branner, Sven
; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5691162o No. 5691162disk of No. 5691162th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,104
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,489
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Agria Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 4180.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-443-104-2

Query Match 34.9%; Score 29; DB 1; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKVMSG 15
| | | | |
Db 2 TYKVYPWG 9

RESULT 2
US-08-238-130-3
; Sequence 3, Application US/08238130
; Patent No. 5702934
; GENERAL INFORMATION:
; APPLICANT: Hastrup, Sven
; APPLICANT: Branner, Sven
; APPLICANT: Jorgensen, Birthe R.
; APPLICANT: Christensen, Iove
; APPLICANT: Jorgensen, Birgitte B.
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; TITLE OF INVENTION: PROCESSES FOR PRODUCING AN ENZYME
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5702934o No. 5702934disk of No. 5702934th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/238,130
; FILING DATE: 04-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 522/93
; FILING DATE: 05-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agnis Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3965.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-238-130-3

Query Match 34.9%; Score 29; DB 1; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKVMSG 15
| | | | |
Db 2 TYKVYPWG 9

RESULT 3
US-08-442-859-2
; Sequence 2, Application US/08442859
; Patent No. 5807729
; GENERAL INFORMATION:
; APPLICANT: Shuster, Jeffrey R.

; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; APPLICANT: Branner, Sven
; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5807729o No. 5807729disk of No. 5807729th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,859
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 252
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,489
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Agnis Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 4180.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-442-859-2

Query Match 34.9%; Score 29; DB 1; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKVMSG 15
| | | | |
Db 2 TYKVYPWG 9

RESULT 4
US-08-398-489-2
; Sequence 2, Application US/08398489
; Patent No. 5843753
; GENERAL INFORMATION:
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; APPLICANT: Branner, Sven
; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5843753o No. 5843753disk of No. 5843753th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,489
FILING DATE: 03-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-398-489-2

Query Match 34.9%; Score 29; DB 2; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKVYMSG 15
| | | | |
Db 2 TYKVYPWG 9

RESULT 5
US-08-894-772-3
Sequence 3, Application US/08894772
Patent No. 5861280
GENERAL INFORMATION:
APPLICANT: Lehmbek, Jan
TITLE OF INVENTION: Host Cell Expressing Reduced Levels
TITLE OF INVENTION: Of A Metalloprotease And Methods Using The Host Cell In Protei
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5861280 No. 5861280disk of No. 5861280th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,772
FILING DATE: 27-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 4300.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-772-3

Query Match 34.9%; Score 29; DB 2; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKVYMSG 15
| | | | |
Db 2 TYKVYPWG 9

RESULT 6
US-09-207-844-3
Sequence 3, Application US/09207844
Patent No. 5968774
GENERAL INFORMATION:
APPLICANT: Lehmbek, Jan
TITLE OF INVENTION: Host Cell Expressing Reduced Levels
TITLE OF INVENTION: Of A Metalloprotease And Methods Using The Host Cell In Protei
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5968774 No. 5968774disk of No. 5968774th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/207,844
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/894,772
FILING DATE: 27-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 4300.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-207-844-3

Query Match 34.9%; Score 29; DB 2; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 TYKVYMSG 15
| | | | |
Db 2 TYKVYPWG 9

RESULT 7
PCT-US95-05534-2
Sequence 2, Application PC/TUS9505534
GENERAL INFORMATION:
APPLICANT:
APPLICANT:

;; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
;; TITLE OF INVENTION: ACTIVITY
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Novo Nordisk of North America, Inc.
;; STREET: 405 Lexington Avenue, 64th floor
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10174-6401
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Tape
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/05534
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/238/108
;; FILING DATE: 04-MAY-1994
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/398,489
;; FILING DATE: 03-MARCH-1995
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Agilis Dr., Cheryl H.
;; REGISTRATION NUMBER: 34,086
;; REFERENCE/DOCKET NUMBER: 4180.204-WO
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-867-0123
;; TELEFAX: 212-878-9655
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
PCT-US95-05534-2

Query Match 34.9%; Score 29; DB 5; Length 14;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8 TYKVYMSG 15
Db 2 TYKVYPWG 9

RESULT 8
US-09-528-200-40
; Sequence 40, Application US/09528200
; Patent No. 6630570
; GENERAL INFORMATION:
; APPLICANT: LICH, KAI
; APPLICANT: BECKER, ANDREAS
; APPLICANT: SEMMLER, WOLFHARD
; APPLICANT: WEIDENMANN, BERTRAM
; APPLICANT: HESSNIUS, CARTSEN
; APPLICANT: VOLKMER-ENGERT, RUDOLF
; APPLICANT: SCHNEIDER-MERGENER, JENS
; APPLICANT: BHARGAVA, SARAH
; TITLE OF INVENTION: SHORT-CHAIN PEPTIDE-DYE CONJUGATES AS CONTRAST MEDIA
; TITLE OF INVENTION: FOR OPTICAL DIAGNOSIS
; FILE REFERENCE: SCH-1731
; CURRENT APPLICATION NUMBER: US/09/528,200
; CURRENT FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: DE 199 17 713.9
; PRIOR FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 196

;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 40
;; LENGTH: 12
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-528-200-40

Query Match 33.7%; Score 28; DB 4; Length 12;
Best Local Similarity 41.7%; Pred. No. 54;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 IAKQMTYKVYMS 14
Db 1 LRQMAYKKYLIN 12

RESULT 9
US-08-413-708B-6
; Sequence 6, Application US/08413708B
; Patent No. 5972883
; GENERAL INFORMATION:
; APPLICANT: GOZES, Iilana
; APPLICANT: FRIDKIN, Matityahu
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS COMPRISING
; TITLE OF INVENTION: VIP, OR ANALOGUES DERIVATIVES AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,708B
; FILING DATE: 30-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/207,671
; FILING DATE: 09-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 105061
; FILING DATE: 16-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: GOZES-3A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-413-708B-6

Query Match 33.7%; Score 28; DB 2; Length 13;
Best Local Similarity 38.5%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 5 KQMTYKVYMSGTV 17
Db 1 KQMAVKKYLAAYL 13

RESULT 10
US-09-528-200-33
; Sequence 33, Application US/09528200
; Patent No. 6630570
; GENERAL INFORMATION:
; APPLICANT: LICHIA, KAI
; APPLICANT: BECKER, ANDREAS
; APPLICANT: SEMMLER, WOLFHARD
; APPLICANT: WEIDENMANN, BERTRAM
; APPLICANT: HESSNIUS, CARSTEN
; APPLICANT: VOLKMER-ENGERT, RUDOLF
; APPLICANT: SCHNEIDER-MERGENER, JENS
; APPLICANT: BHARGAVA, SARAH
; TITLE OF INVENTION: SHORT-CHAIN PEPTIDE-DYE CONJUGATES AS CONTRAST MEDIA
; FILE REFERENCE: SCH-1731
; CURRENT APPLICATION NUMBER: US/09/528,200
; PRIOR FILING DATE: 2000-03-17
; PRIOR FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-528-200-33

Query Match 33.7%; Score 28; DB 4; Length 13;
Best Local Similarity 41.7%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 IAKQMTYKVYMS 14
: ||| |::
Db 1 LRKQMAVKKYLN 12

RESULT 11
US-09-528-200-39
; Sequence 39, Application US/09528200
; Patent No. 6630570
; GENERAL INFORMATION:
; APPLICANT: LICHIA, KAI
; APPLICANT: BECKER, ANDREAS
; APPLICANT: SEMMLER, WOLFHARD
; APPLICANT: WEIDENMANN, BERTRAM
; APPLICANT: HESSNIUS, CARSTEN
; APPLICANT: VOLKMER-ENGERT, RUDOLF
; APPLICANT: SCHNEIDER-MERGENER, JENS
; APPLICANT: BHARGAVA, SARAH
; TITLE OF INVENTION: SHORT-CHAIN PEPTIDE-DYE CONJUGATES AS CONTRAST MEDIA
; FILE REFERENCE: SCH-1731
; CURRENT APPLICATION NUMBER: US/09/528,200
; PRIOR FILING DATE: 2000-03-17
; PRIOR FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-528-200-39

Query Match 33.7%; Score 28; DB 4; Length 13;
Best Local Similarity 41.7%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 IAKQMTYKVYMS 14
: ||| |::
Db 2 LRKQMAVKKYLN 13

RESULT 12
US-09-528-200-26
; Sequence 26, Application US/09528200
; Patent No. 6630570
; GENERAL INFORMATION:
; APPLICANT: LICHIA, KAI
; APPLICANT: BECKER, ANDREAS
; APPLICANT: SEMMLER, WOLFHARD
; APPLICANT: WEIDENMANN, BERTRAM
; APPLICANT: HESSNIUS, CARSTEN
; APPLICANT: VOLKMER-ENGERT, RUDOLF
; APPLICANT: SCHNEIDER-MERGENER, JENS
; APPLICANT: BHARGAVA, SARAH
; TITLE OF INVENTION: SHORT-CHAIN PEPTIDE-DYE CONJUGATES AS CONTRAST MEDIA
; FILE REFERENCE: SCH-1731
; CURRENT APPLICATION NUMBER: US/09/528,200
; PRIOR FILING DATE: 2000-03-17
; PRIOR FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-528-200-26

Query Match 33.7%; Score 28; DB 4; Length 14;
Best Local Similarity 41.7%; Pred. No. 65;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 IAKQMTYKVYMS 14
: ||| |::
Db 1 LRKQMAVKKYLN 12

RESULT 13
US-09-528-200-32
; Sequence 32, Application US/09528200
; Patent No. 6630570
; GENERAL INFORMATION:
; APPLICANT: LICHIA, KAI
; APPLICANT: BECKER, ANDREAS
; APPLICANT: SEMMLER, WOLFHARD
; APPLICANT: WEIDENMANN, BERTRAM
; APPLICANT: HESSNIUS, CARSTEN
; APPLICANT: VOLKMER-ENGERT, RUDOLF
; APPLICANT: SCHNEIDER-MERGENER, JENS
; APPLICANT: BHARGAVA, SARAH
; TITLE OF INVENTION: SHORT-CHAIN PEPTIDE-DYE CONJUGATES AS CONTRAST MEDIA
; FILE REFERENCE: SCH-1731
; CURRENT APPLICATION NUMBER: US/09/528,200
; PRIOR FILING DATE: 2000-03-17
; PRIOR FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-528-200-32

Query Match 33.7%; Score 28; DB 4; Length 14;
Best Local Similarity 41.7%; Pred. No. 65;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 IAKQMTYKVYMS 14
: ||| | : :
Db 2 LRQMAVKKYL 13

RESULT 14
US-08-486-721A-16
; Sequence 16, Application US/08486721A
; Patent No. 5739025
; GENERAL INFORMATION:
; APPLICANT: Fukazawa, Chikafusa
; TITLE OF INVENTION: Method of Producing
; TITLE OF INVENTION: Asparaginyl Endoprotease
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frischauf, Holtz, Goodman, Langer & Chick
; STREET: 767 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10017-2023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3+ inch, 1.4 mb
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,721A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/091,991
; FILING DATE: 12-JULY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard S.
; REGISTRATION NUMBER: 28,180
; REFERENCE/DOCKET NUMBER: 930587/HG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 319-4900
; TELEFAX: (212) 319-5101
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:

; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
US-08-486-721A-16
Query Match 33.7%; Score 28; DB 1; Length 15;
Best Local Similarity 41.7%; Pred. No. 70;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 3 IAKQMTYKVYMS 14
: ||| | : :
Db 4 LRQMAVKKYL 15

RESULT 15
US-09-528-200-19
; Sequence 19, Application US/09528200
; Patent No. 6630570
; GENERAL INFORMATION:
; APPLICANT: LICHA, KAI
; APPLICANT: BECKER, ANDREAS
; APPLICANT: SEMMLER, WOLFGANG
; APPLICANT: WEIDENMANN, BERTRAM
; APPLICANT: HESSNIUS, CARSTEN
; APPLICANT: VOLKMER-ENGERT, RUDOLF
; APPLICANT: SCHNEIDER-MERGEN, JENS
; APPLICANT: BHARGAVA, SARAH
; TITLE OF INVENTION: SHORT-CHAIN PEPTIDE-DYE CONJUGATES AS CONTRAST MEDIA
; TITLE OF INVENTION: FOR OPTICAL DIAGNOSIS
; FILE REFERENCE: SCH-1731
; CURRENT APPLICATION NUMBER: US/09/528,200
; CURRENT FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: DE 199 17 713.9
; PRIOR FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 19
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-528-200-19
Query Match 33.7%; Score 28; DB 4; Length 15;
Best Local Similarity 41.7%; Pred. No. 70;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 3 IAKQMTYKVYMS 14
: ||| | : :
Db 1 LRQMAVKKYL 12

Search completed: August 12, 2004, 07:06:30
Job time : 41 secs